A PATIENT-CENTERED APPROACH TO BENEFIT-HARM ASSESSMENT IN TREATMENT DECISION-MAKING

by
Tsung Yu

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
October, 2014

ABSTRACT

Several structured and quantitative approaches have been developed to determine the benefit-harm balance of medical treatments in specific populations. These approaches summarize the key factors of a benefit-harm assessment that may include baseline outcome risks, treatment effects and relative importance for outcomes. The overarching goal of this dissertation was to develop a patient-centered approach to benefit-harm assessment in treatment decision-making.

In part one of this dissertation, we reviewed the labels and medical reviews of 58 drugs approved by the US Food and Drug Administration for four diseases to learn how they dealt with surrogate outcomes when they were assessing the benefits and harms of drugs. Most drugs for chronic obstructive pulmonary disease, diabetes, and glaucoma were approved based only on surrogates but for osteoporosis, most drugs were also approved for patient-centered outcomes. The rationale for using surrogates was not often discussed (11 out of the 43 drug approvals based only on surrogates, 26%) in medical reviews. We accordingly proposed a framework for use of surrogate outcomes in doing patient-centered benefit-harm assessments.

In part two of this dissertation, we conducted a survey of patients with non-infectious uveitis to elicit their preferences for six treatment outcomes associated with corticosteroid therapy. Eighty-two patients in the Multicenter Uveitis Steroid Treatment (MUST) Trial Follow-up Study and 100 patients treated at two academic medical centers

(Johns Hopkins University and University of Pennsylvania) completed the best-worst scaling tasks, in which they repeatedly selected the most and least worrying from a list of three outcomes. Results showed that participants were more likely to select vision not meeting the requirement for driving, glaucoma, and needing eye surgery as the most worrying outcomes as compared against needing medicine for high blood pressure/cholesterol, cataracts or infection (e.g., sinusitis).

In part three of this dissertation, we conducted a quantitative benefit-harm assessment using data from the MUST trial that compared corticosteroid implant versus systemic corticosteroids in non-infectious intermediate, posterior, and panuveitis. We calculated benefit-harm metrics to reflect the benefit-harm balance that took into account the treatment effects on different patient-centered outcomes and the patient preferences for these outcomes (derived from the preference-elicitation survey). The benefit-harm metrics at 6, 12, 18, and 24 months follow-up were all negative and the probabilities of the metric being positive were small or 0%. This implied that implant therapy had a worse benefit-harm balance than systemic therapy.

In summary, using an example of corticosteroid therapy for treating non-infectious uveitis, we demonstrated a patient-centered approach to benefit-harm assessment where we focused on patient-centered outcomes and incorporated patient preferences to estimate treatment benefit-harm balance. Our approach can be applied in the future to different diseases and settings and help make evidence- and preference-based treatment decisions.

COMMITTEE OF FINAL THESIS READERS

Advisors:

Milo A. Puhan, MD, PhD

Epidemiology, Biostatistics and Prevention Institute, University of Zurich

Janet T. Holbrook, PhD

Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health

Readers:

Thomas A. Louis, PhD

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health

Cynthia M. Boyd, MD

Geriatric Medicine, Johns Hopkins School of Medicine

Jodi B. Segal, MD

General Internal Medicine, Johns Hopkins School of Medicine

Tianjing Li, MD, PhD

Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health

Antonio J. Trujillo, PhD

Department of International Health, Johns Hopkins Bloomberg School of Public Health

ACKNOWLEDGEMENTS

I would not complete my dissertation without many people's help. My most sincere thanks go to my advisor, Dr. Milo Puhan, who dedicated enormous time to my training. He is a passionate researcher and actively involves me in his work. Although he left Hopkins two years ago, he is always responsive to any emails that I send to him. It is all his encouragement that gives me confidence in traveling on this "bumpy" road to PhD.

Many important individuals helped me a lot on my dissertation. Dr. Janet Holbrook provided me strong support for making it possible to do patient survey in the MUST FS and at other clinics. She gave me many insights on data collection and analysis and I learned so much when actually working on primary studies. Researchers in the MUST Trial research group or at University of Pennsylvania, including Nancy Prusakowski, Tonetta Fitzgerald, Alyce Burke, Mark van Natta, John Kempen, Douglas Jabs, Jennifer Thorne and others all gave me help along the way. Thanks also go to Kevin Fain, Yea-Jen Hsu, Cynthia Boyd, Thomas Louis, Kevin Frick and all examiners and readers for reading and commenting on my dissertation.

During my studies, I am fortunate as well to work with Drs. Kay Dickersin and Tianjing Li and colleagues in the Cochrane Eyes and Vision Group. I would for sure remember all the conferences we have been to and the good trips attached. All the classmates and friends I met in Baltimore made my life in the past few years much more colorful. I hope to keep this friendship in the future since it means so much to me!

Finally, my special thanks go to my family, especially my parents, for their unconditional love. Without their support, I would not be able to come to Hopkins for this degree. 謝謝 (Thank you)!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

Making treatment decisions is difficult. The goal of treatments is to provide benefits, but they can come with harms. For example, aspirin is prescribed to patients to prevent cardiovascular diseases, but there is a risk of gastrointestinal and brain bleeds when taking aspirin.[1] Thus, it is not clear in whom aspirin would be beneficial overall so the benefits (lowered risk of cardiovascular diseases) outweigh the harms (increased risk of bleeds).[2] Patients and clinicians alike often face the uncertainty if a treatment is more beneficial than harmful to them or to their patients, or if it is worthwhile to bear the harms of a treatment in return for its benefits. Ideally, patients and clinicians involved should be well-informed of the available treatment options and their associated benefits and harms when they are faced with making treatment decisions.[3, 4]

Not only do the patients and clinicians face the challenge of weighing benefits and harms in making treatment decisions, regulatory bodies also face the same challenge since their key task is to ensure that new medical products are safe and effective for use in the population.[5, 6] This requires the regulators to carefully assess the benefits and harms of drugs before making regulatory decisions. Any decision made by the regulatory bodies is dependent upon many factors including scientific results, uncertainties and subjective judgments.[6] The US Food and Drug Administration (FDA) has been criticized previously for not having enough clarity and transparency about their drug review process.[6] To address this issue, they made several commitments to developing and implementing structured approaches to benefit-harm assessment in making drug regulatory decisions.[6]

Previous research has proposed a number of approaches to benefit-harm assessment.[7-10] It can either be done *qualitatively* by informal synthesis of benefits and harms or instead *quantitatively* by building models that incorporate all data elements influencing the benefit-harm balance.[7-10] We argue here that using a quantitative approach to benefit-harm assessment is more explicit, consistent, and transparent than a qualitative assessment because all elements that go into the analysis are clearly laid out and because the impact of varying these elements can easily be assessed.[9] Thereby, it can provide a basis for decision-making and improve the communication between different stakeholders (e.g., patients, clinicians, and policy-makers).

Critical elements for benefit-harm assessment

Many quantitative approaches to benefit-harm assessment were developed to synthesize three critical data elements in estimating the benefit-harm balance: treatment effects on outcomes, baseline outcome risks, and relative importance of outcomes.[9] Treatments are often evaluated in randomized clinical trials to learn their effects on multiple benefit and harm outcomes. When attempting to apply these treatment effects to a patient population, we need also to account for the baseline outcome risks, which are the background risks for outcomes if patients were not treated. As in the example of aspirin, prescribing aspirin to elder patients (who are at a higher risk for gastrointestinal bleeds) is likely to cause more gastrointestinal bleeds events than to younger patients (who are at a lower risk for gastrointestinal bleeds).[2] Moreover, different outcomes can

mean very differently to patients.[11] In conducting benefit-harm assessment, investigators often assign weights (relative importance) to outcomes when comparing different outcomes against each other. These weights can be informed by patient preferences but may vary depending on whose perspectives are of interest (e.g., patients', clinicians', or policy-makers' perspective).[8]

Here we briefly give one example of a comprehensive benefit-harm assessment in which the investigators used absolute risks (number of events prevented or in excess) to assess the benefits and harms of Tamoxifen in preventing breast cancer (see **Appendix 1-1**).[12]

Puhan and his colleagues have proposed a framework to organize and select different quantitative approaches to conducting benefit-harm assessment, which was based on the number of benefit and harm outcomes and whether a comparison metric is used.[9] Some approaches to benefit-harm assessment consider only a single outcome for benefits and a single outcome for harms and some approaches consider more than one outcome for benefits or harms. The metrics being used to compare the benefit and harm outcomes are different across approaches as well.[9] For example, some people use number needed to treat/number needed to harm (NNT/NNH) ratio[13, 14] and some use quality-adjusted life years (QALYs)[15] to synthesize different elements of a benefit-harm assessment. A number of challenges that investigators may encounter while conducting the assessment were identified in previous reviews[7-10] and are summarized

in the **Table 1-1**. Two of the challenges are of main interest in this dissertation: selecting outcomes for benefit-harm assessment and assessing relative importance for treatment outcomes.

Selecting outcomes for benefit-harm assessment

Selecting outcomes is one key step for a benefit-harm assessment because we want to capture the treatment effects that are meaningful to patients.[3, 4] We advocate that investigators should first consider patient-centered outcomes, instead of surrogate outcomes, for conducting their benefit-harm assessments. Surrogate outcomes, defined by Temple[16] as "a laboratory measurement or a physical sign used as a substitute for a clinically meaningful outcome that measures directly how a patient feels, functions or survives", are commonly used in clinical trials to demonstrate treatment effects.[17] Low bone mineral density, for instance, is a surrogate outcome that is not immediately felt by patients or does not immediately affect the functioning of patients. Instead, it is a risk factor for fractures that represent a patient-centered outcome immediately linked to how patients function in everyday life.[18]

Surrogate outcomes may correlate well with the patient-centered outcomes that affect the health of patients but may not fully capture the effect of treatments.[19] Examples such as CD4 cell count and progression of AIDS,[20] tumor shrinkage and survival of cancer patients[21] and arrhythmias and sudden cardiac death[22] show that

use of surrogate outcomes in replacement of patient-centered outcomes may be misleading about the treatment effect.[23] Ideally, treatment effects on surrogate outcomes should have been shown to predict treatment effects on patient-centered outcomes before surrogate outcomes can be used in clinical trials.[24] For benefit-harm assessment, however, it is currently unclear how to include surrogate outcomes in the assessment. Moreover, if treatment effects are only evaluated for surrogate outcomes, how can investigators proceed?

Assessing relative importance for treatment outcomes

Another key step in doing a benefit-harm assessment is to properly weigh the relative importance of treatment outcomes, which can be informed by patient preferences.[25] Patients value different health outcomes at different importance. An individual's preference reflects the degree of their subjective satisfaction, distress or desirability for a given health outcome.[26, 27] The trade-offs between different benefits and harms of treatments is thus largely influenced by how patients place the relative importance on each outcome.[28] In the United States, the newly established Patient-Centered Outcomes Research Institute aims to conduct research that engages more of the patients' perspective, interests, and values with the hope that patients can make informed healthcare decisions that lead them to better outcomes.[3]

Different methods have been used to assess patient preferences, e.g., ranking or rating, standard gamble (SG), time trade-off (TTO), and visual analogue scale (VAS).[29, 30] The SG is an instrument based on a decision-making model that involves uncertainty, where participants are asked to choose either a gamble between perfect health and death or living in a given health state (an intermediate health state between perfect health and death).[30] The TTO, developed as an alternative to SG, is an instrument in which participants are asked to choose between two options: living in a certain health state for time t followed by death or living in perfect health for time x ($x < t$) followed by death.[30] Both SG and TTO are conceptually nice ways to assess patient preferences but are challenging for respondents and thus may preclude from obtaining reliable data.[31]

Another method to obtain patient preferences is the VAS, where participants are asked to assign their preference for a health state on a linear rating scale anchored by perfect health and death.[30] One limitation of VAS is that, unlike SG, it does not reflect conditions of uncertainty under which medical decisions are usually made. It also does not involve "trade-offs" while methods such as SG and TTO ask participants about trade-offs between different health states.[30] These preference-elicitation approaches developed in health economics are helpful for us to understand how patients make trade-offs between efficacy and safety and to better use evidence from clinical trials or epidemiologic studies to make treatment decisions.[32] However, the challenging tasks of these approaches may sometimes question the reliability of the obtained preferences.

Best-worst scaling (BWS) method may address this operational challenge. It is another type of preference-elicitation approach that was designed in marketing research and has gained popularity in recent years in healthcare research.[33-36] In BWS, participants are presented with different sets of "objects", and then they are asked to select the best and the worst objects in each set. The importance of an object relative to others (preferences) can be inferred statistically from a series of stated choices that participants make in the BWS tasks (see **Appendix 1-2** for details). BWS has been shown to be a less cognitively demanding approach to eliciting preferences; thus, it has a great potential for use in the clinical settings.[33]

The clinical example: uveitis and corticosteroid therapy

In this dissertation, we will illustrate our approach to benefit-harm assessment with the example of corticosteroid therapies for treating non-infectious uveitis.

Uveitis includes a wide range of clinical conditions where inflammation affects components of the uvea, i.e., the iris, ciliary body, and choroid. The classification system recommended by the International Uveitis Study Group and the Standardization of Uveitis Nomenclature working group is based on the anatomical location.[37, 38] They categorized uveitis into four types: anterior, intermediate, posterior, and panuveitis. Uveitis can also be categorized on the basis of its onset, duration, and course. Causes of

uveitis include microbial infection, trauma, and autoimmunity but the cause remains unknown in many cases.[39]

Non-infectious uveitis can cause vision loss and accounts for around 10% of blindness in the United States.[40] Adults aged from 20 to 60 years are more often affected by uveitis than other age groups.[41] Thus, compared to other age-related eye diseases such as cataracts, glaucoma, and age-related macular degeneration, uveitis has a disproportionately high socioeconomic impact because it affects mainly the working population.[42] The goal of treating uveitis is to eliminate the inflammation within the eye to preserve patients' vision while minimizing the potential harms associated with the treatments.

Corticosteroid usage is the primary treatment for uveitis. This group of medications has strong anti-inflammatory and immunosuppressive effects. The modes of delivery of corticosteroids include topical, periocular, intravitreal, and systemic routes. Topical corticosteroids are used mainly for anterior uveitis because they cannot penetrate the posterior segment well to achieve the necessary concentrations for intermediate or posterior uveitis.[43] Periocular corticosteroid injections are effective in treating intermediate or posterior uveitis and are particularly useful in patients with unilateral or asymmetric diseases.[44]

Corticosteroids can also be delivered intravitreally by injection of corticosteroids into the vitreous body[45] or by implantation of sustained-release delivery vehicles, e.g., Ozurdex® (Allergan, Inc., Irvine, CA, USA) and Retisert® (Bausch & Lomb, Rochester, NY, USA).[46] The sustained-release implant releases corticosteroids over a period of time (months to years) and has been found to be effective in non-infectious uveitis patients.[47] Systemic (oral) corticosteroids are often used in patients with intermediate and posterior uveitis that is bilateral or related to systemic diseases, or in patients poorly responsive to topical and periocular corticosteroids.[47] When corticosteroids are used in a local (topical, periocular, and intravitreal) way, the systemic side effects associated oral corticosteroid therapy can be limited. However, the risks of developing cataracts and elevated intraocular pressure or glaucoma are much higher.[48]

The Multicenter Uveitis Steroid Treatment Trial

In 2004, the National Eye Institute funded the study of the Multicenter Uveitis Steroid Treatment (MUST) Trial,[49, 50] which is a comparative effectiveness clinical trial comparing local fluocinolone acetonide implant therapy with systemic corticosteroids plus immunosuppression when indicated (current standard of care) for non-infectious intermediate, posterior, and panuveitis. The aim of the MUST Trial was to determine if the implant therapy is better than the systemic therapy in terms of the effectiveness and safety. Two hundreds and fifty-five patients for whom systemic

corticosteroids were indicated were randomized. Change in best-corrected visual acuity from baseline to 24 months was defined as the primary outcome in the trial.

At 24 months follow-up, results of the MUST Trial showed that the improvement in visual acuity between the implant therapy group and the systemic therapy group was not statistically significantly different (6.0 versus 3.2 letters, p=0.16). The implant therapy group showed a better control of uveitis activity during follow-up (88% versus 71% controlled at 24 months, p=0.001). However, the implant group had a higher risk of developing cataracts (91% versus 45 %; hazard ratio=4.1, p<0.0001) or glaucoma (17% versus 4.0%; hazard ratio=4.2, p=0.0008). The risk of adverse systemic outcomes (including systemic infections, hypertension, hyperlipidemia, etc.) was generally lower in the implant therapy group although most effects were not statistically significant. The investigators of the MUST Trial concluded that "the specific advantages and disadvantages identified should dictate selection between the alternative treatments in consideration of individual patients' particular circumstances."[50] This indicates that there is still much uncertainty about the use of fluocinolone acetonide implant for patients with non-infectious uveitis given its associated harms.

Findings from the MUST Trial therefore provided us with a great opportunity to investigate new approaches to benefit-harm assessment, because each treatment strategy was associated with treatment outcomes that occurred at different frequencies and that had various importance to patients. It is perhaps impossible or challenging, without using

a structured and quantitative approach, to properly synthesize all these elements to learn the treatment benefit-harm balance. Ideally, such an approach will clearly define the treatment outcomes of interest, assess patient preferences (relative importance) for outcomes and synthesize the treatment effects on outcomes with their relative importance to facilitate the comparison between treatments and thereby help make evidence- and preference- based decisions.[9]

Overview of the dissertation

The overarching goal of this dissertation was to develop a patient-centered approach to benefit-harm assessment that focuses patient-centered outcomes and that considers patient preferences. Using a clinical example, we quantitatively assessed the benefits and harms of two treatment strategies in non-infectious uveitis to learn their benefit-harm balance. At the same time we demonstrated how incorporating patients' perspectives can be achieved in medical research. **Chapter 2** is a survey of the drug approvals by the US FDA to examine how surrogate outcomes were dealt with when they were assessing the benefits and harms of a drug, and accordingly we developed a framework for selecting outcomes in benefit-harm assessments. **Chapter 3** is a patient preferences study in which we used BWS approach to elicit outcome preferences by surveying patients with non-infectious uveitis. We obtained the relative importance of the six treatment outcomes, as perceived by patients. **Chapter 4** is a quantitative benefit-harm assessment that is based on a comparative effectiveness trial (the MUST Trial)

comparing corticosteroid implant versus systemic corticosteroids in non-infectious uveitis. We focused on patient-centered outcomes and incorporated patient preferences in our assessment. Finally, in **Chapter 5**, we summarized our findings of the dissertation and discussed the limitations and implications of our research.

REFERENCES

1. US Preventive Services Task Force. Aspirin for the prevention of cardiovascular disease: U.S. preventive services task force recommendation statement. Ann Intern Med. 2009 Mar 17;150(6):396-404.
2. Puhan MA, Singh S, Weiss CO, Varadhan R, Sharma R, Boyd CM. Evaluation of the Benefits and Harms of Aspirin for Primary Prevention of Cardiovascular Events: A Comparison of Quantitative Approaches [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013 Nov.
3. Patient-Centered Outcomes Research Institute. PCORI Funding Announcement: Pilot Projects Grants. Washington, DC: Patient-Centered Outcomes Research Institute; http://www.pcori.org/research-we-support/pcor/. 2013 (last accessed on June 3, 2014).
4. Barry MJ, Edgman-Levitan S. Shared decision making--pinnacle of patient-centered care. N Engl J Med. 2012 Mar 1;366(9):780-1.
5. Drug Benefits and Risks: International Textbook of Clinical Pharmacology, revised 2nd edition. Edited by C.J. van Baxtel, B. Santoso, and J.R. Edwards. IOS Press and Upsala Monitoring Centre, 2008.
6. The Food and Drug Administration. Structured approach to benefit-risk assessment in drug regulatory decision-making. Draft PDUFA V implementation plan - February 2013 fiscal years 2013-2017. http://www.fda.gov/downloads/ForIndustry/UserFees/PrescriptionDrugUserFee/UCM329758.pdf. 2013 (last accessed on January 17, 2014).

7. Guo JJ, Pandey S, Doyle J, Bian B, Lis Y, Raisch DW. A review of quantitative risk-benefit methodologies for assessing drug safety and efficacy-report of the ISPOR risk-benefit management working group. Value Health. 2010;13(5):657-66.
8. Mt-Isa S, Hallgreen CE, Wang N, Callréus T, Genov G, Hirsch I, Hobbiger SF, Hockley KS, Luciani D, Phillips LD, Quartey G, Sarac SB, Stoeckert I, Tzoulaki I, Micaleff A, Ashby D; IMI-PROTECT benefit-risk participants. Balancing benefit and risk of medicines: a systematic review and classification of available methodologies. Pharmacoepidemiol Drug Saf. 2014 Jul;23(7):667-78.
9. Puhan MA, Singh S, Weiss CO, Varadhan R, Boyd CM. A framework for organizing and selecting quantitative approaches for benefit-harm assessment. BMC Med Res Methodol. 2012;12:173,2288-12-173.
10. The PROTECT Consortium. Review of methodologies for benefit and risk assessment of medication. http://www.imi-protect.eu/documents/ShahruletalReviewofmethodologiesforbenefitandriskassessmentofmedicationMay2013.pdf. 2012 (last accessed on April 12, 2014).
11. Eraker SA, Sox HC Jr. Assessment of patients' preferences for therapeutic outcomes. Med Decis Making. 1981;1(1):29-39.
12. Gail MH, Costantino JP, Bryant J, Croyle R, Freedman L, Helzlsouer K, Vogel V. Weighing the risks and benefits of tamoxifen treatment for preventing breast cancer. J Natl Cancer Inst. 1999;91(21):1829-46.
13. Laupacis A, Sackett DL, Roberts RS: An assessment of clinically useful measures of the consequences of treatment. N Engl J Med 1988, 318(26):1728–1733.

14. Osiri M, Suarez-Almazor ME, Wells GA, Robinson V, Tugwell P. Number needed to treat (NNT): implication in rheumatology clinical practice. Ann Rheum Dis. 2003 Apr;62(4):316-21.
15. Garrison LP Jr, Towse A, Bresnahan BW: Assessing a structured, quantitative health outcomes approach to drug risk-benefit analysis. Health Aff (Millwood) 2007, 26(3):684–695.
16. Temple RJ. A regulatory authority's opinion about surrogate endpoints. Clinical Measurement in Drug Evaluation. Edited by Nimmo WS, Tucker GT. New York: Wiley; 1995.
17. Biomarkers Definition Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther. 2001;69:89-95.
18. The Food and Drug Administration. Guidelines for preclinical and clinical evaluation of agents used in the prevention or treatment of postmenopausal osteoporosis. http://www.fda.gov/downloads/ScienceResearch/SpecialTopics/WomensHealthResearch/UCM131206.pdf. 1994 (last accessed on June 16, 2014).
19. Fleming TR, Powers JH. Biomarkers and surrogate endpoints in clinical trials. Stat Med. 2012 Nov 10;31(25):2973-84.
20. Sande MA, Carpenter CC, Cobbs CG, Holmes KK, Sanford JD. Anti-retroviral therapy for adult HIV-infected patients. Recommendations from a state-of-the-art conference. National Institute of Allergy and Infectious Diseases State-of-the-Art

Panel on Anti-Retroviral Therapy for Adult HIV-Infected Patients. JAMA. 1993; 270:2583-9.

21. Modulation of fluorouracil by leucovorin in patients with advanced colorectal cancer: evidence in terms of response rate. Advanced Colorectal Cancer Meta-Analysis Project. J Clin Oncol. 1992; 10:896-903.
22. Effect of the antiarrhythmic agent moricizine on survival after myocardial infarction. The Cardiac Arrhythmia Suppression Trial II Investigators. N Engl J Med. 1992; 327:227-33.
23. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? Ann Intern Med. 1996;125(7):605-13.
24. International Conference on Harmonization. ICH harmonized tripartite guideline. Statistical principles for clinical trials - E9. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf. 1998 (last accessed on June 16, 2014).
25. van Til JA, Ijzerman MJ. Why should regulators consider using patient preferences in benefit-risk assessment? Pharmacoeconomics. 2014 Jan;32(1):1-4.
26. Wennberg JE. Unwarranted variations in healthcare delivery: implications for academic medical centres. BMJ. 2002 Oct 26;325(7370):961-4.
27. MacLean S, Mulla S, Akl EA, Jankowski M, Vandvik PO, Ebrahim S, McLeod S, Bhatnagar N, Guyatt GH; American College of Chest Physicians. Patient values and preferences in decision making for antithrombotic therapy: a systematic review: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American

College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest. 2012;141(2 Suppl):e1S-23S.

28. Blinman P, King M, Norman R, Viney R, Stockler MR. Preferences for cancer treatments: an overview of methods and applications in oncology. Ann Oncol. 2012 May;23(5):1104-10.
29. Kymes SM, Lee BS. Preference-based quality of life measures in people with visual impairment. Optom Vis Sci. 2007;84(8):809-16.
30. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the economic evaluation of health care programmes. Oxford university press, 2005.
31. Puhan MA, Guyatt GH, Montori VM, Bhandari M, Devereaux PJ, Griffith L, Goldstein R, Schünemann HJ.J The standard gamble demonstrated lower reliability than the feeling thermometer. Clin Epidemiol. 2005 May;58(5):458-65.
32. Briggs AH, Levy AR. Pharmacoeconomics and pharmacoepidemiology: curious bedfellows or a match made in heaven? Pharmacoeconomics. 2006;24(11):1079-86.
33. Flynn TN. Valuing citizen and patient preferences in health: recent developments in three types of best-worst scaling. Expert Rev Pharmacoecon Outcomes Res. 2010 Jun;10(3):259-67.
34. Flynn TN, Louviere JJ, Peters TJ, Coast J. Best--worst scaling: What it can do for health care research and how to do it. J Health Econ. 2007 Jan;26(1):171-89.
35. Gallego G, Bridges JF, Flynn T, Blauvelt BM, Niessen LW. Using best-worst

scaling in horizon scanning for hepatocellular carcinoma technologies. Int J Technol Assess Health Care. 2012 Jul;28(3):339-46.

36. Louviere JJ, Lings I, Islam T, Gudergan S, Flynn T. An introduction to the application of (case 1) best–worst scaling in marketing research. Int J Res Mark. 2013;30(3):292–303.
37. Deschenes J, Murray PI, Rao NA, Nussenblatt RB; International Uveitis Study Group. International Uveitis Study Group (IUSG): clinical classification of uveitis. Ocul Immunol Inflamm. 2008;16(1):1-2.
38. Jabs DA, Nussenblatt RB, Rosenbaum JT; Standardization of Uveitis Nomenclature (SUN) Working Group. Standardization of uveitis nomenclature for reporting clinical data. Results of the First International Workshop. Am J Ophthalmol. 2005;140(3):509-16.
39. McCannel CA, Holland GN, Helm CJ, Cornell PJ, Winston JV, Rimmer TG. Causes of uveitis in the general practice of ophthalmology. UCLA Community-Based Uveitis Study Group. Am J Ophthalmol. 1996;121(1):35-46.
40. Nussenblatt RB. The natural history of uveitis. Int Ophthalmol. 1990 Oct;14(5-6):303-8.
41. Rothova A, Suttorp-van Schulten MS, Frits Treffers W, Kijlstra A. Causes and frequency of blindness in patients with intraocular inflammatory disease. Br J Ophthalmol. 1996;80(4):332-6.
42. Durrani OM, Meads CA, Murray PI. Uveitis: a potentially blinding disease. Ophthalmologica. 2004;218(4):223-36.

43. Yeh S, Faia LJ, Nussenblatt RB. Advances in the diagnosis and immunotherapy for ocular inflammatory disease. Semin Immunopathol. 2008;30(2):145-64.

44. Ferrante P, Ramsey A, Bunce C, Lightman S. Clinical trial to compare efficacy and side-effects of injection of posterior sub-Tenon triamcinolone versus orbital floor methylprednisolone in the management of posterior uveitis. Clin Experiment Ophthalmol. 2004;32(6):563-8.

45. Inoue M, Takeda K, Morita K, Yamada M, Tanigawara Y, Oguchi Y. Vitreous concentrations of triamcinolone acetonide in human eyes after intravitreal or subtenon injection. Am J Ophthalmol. 2004;138(6):1046-8.

46. Siddique SS, Shah R, Suelves AM, Foster CS. Road to remission: a comprehensive review of therapy in uveitis. Expert Opin Investig Drugs. 2011;20(11):1497-515.

47. Wakefield D, McCluskey P. Systemic immunosuppression in the treatment of posterior uveitis. Int Ophthalmol Clin. 1995;35(3):107-22.

48. Galor A, Margolis R, Brasil OM, Perez VL, Kaiser PK, Sears JE, Lowder CY, Smith SD. Adverse events after intravitreal triamcinolone in patients with and without uveitis. Ophthalmology. 2007;114(10):1912-8.

49. Multicenter Uveitis Steroid Treatment Trial Research Group, Kempen JH, Altaweel MM, Holbrook JT, Jabs DA, Sugar EA. The multicenter uveitis steroid treatment trial: rationale, design, and baseline characteristics. Am J Ophthalmol. 2010 Apr;149(4):550-561.

50. Multicenter Uveitis Steroid Treatment (MUST) Trial Research Group, Kempen

JH, Altaweel MM, Holbrook JT, Jabs DA, Louis TA, Sugar EA, Thorne JE. Randomized comparison of systemic anti-inflammatory therapy versus fluocinolone acetonide implant for intermediate, posterior, and panuveitis: the multicenter uveitis steroid treatment trial. Ophthalmology. 2011;118(10):1916-26.

Table 1-1. Challenges for conducting benefit-harm assessment

Challenges	
Defining decision-making context	Benefits and harms vary depending on the characteristics of the disease, population, intervention and comparison.
Selecting outcomes	Surrogate outcomes are often reported in some diseases and they may not fully substitute for patient-centered outcomes. Suboptimal reporting on harm outcomes. A single benefit and a single harm outcome, or multiple benefit and harm outcomes need to be considered in the analysis.
Defining time frames	Benefits and harms occur at various time points (e.g., early benefits but late harms).
Assessing quality of the evidence	The evidence on benefits and harms is obtained from studies with various qualities.
Obtaining data	Individual patient data are not accessible.
Assessing relative importance for outcomes	The weight (relative importance) of outcomes is considered differently by different stakeholders.
Choosing comparison metrics	Whether to put the benefits and harms on the same metric and the metric chosen affect the benefit-harm comparison.

Appendix 1-1. An example of quantitative benefit-harm assessment: Tamoxifen for preventing breast cancer

Gail and his colleagues in the National Cancer Institute developed an approach to benefit-harm assessment that can deal with the situation when treatments are associated with multiple outcomes.[1] Use of Tamoxifen in women can reduce the risk of breast cancer and bone fractures; however, it also increases the risk of endometrial cancer, stroke, and pulmonary embolism. In Gail's quantitative assessment, they first estimated in a cohort of women, the risk of different outcomes with treatment and without treatment using evidence from observational studies (baseline outcome risks) combined with clinical trials (treatment effects). They calculated, both with and without Tamoxifen treatment, the number of different outcomes expected per 10000 women over five years, stratified by different age and race. Then, they computed treatment "benefits" as number of events prevented and "harms" as number of events in excess. For instance, for white women aged 40-49 years with 2.0% 5-year risk of breast cancer, with the use of Tamoxifen, 97 cases of breast cancer will be prevented but 16 cases of endometrial cancer will be in excess per 10000 women over five years. Finally, in order to put multiple outcomes on the same metric, they assigned a weight to each outcome according to its clinical importance (e.g., 1 for life threatening, 0.5 for severe, and 0 for others), and presented the results as a net number of events prevented or in excess per 10000 women.

[1] Gail MH, Costantino JP, Bryant J, Croyle R, Freedman L, Helzlsouer K, Vogel V. Weighing the risks and benefits of tamoxifen treatment for preventing breast cancer. J Natl Cancer Inst. 1999;91(21):1829-46.

By conducting the benefit-harm assessment, they were able to identify the women's risk profiles for which Tamoxifen is most beneficial.

Appendix 1-2. Best-worst scaling

Best-worst scaling (BWS) is an approach to eliciting consumer preferences and includes 3 different types (cases): Case 1 (object case), Case 2 (profile case), and Case 3 (multi-profile case).[2] Case 1 BWS is the simplest form, in which individuals are asked each time to choose the top and bottom ranked ones ("best" and "worst") from a list of "objects". In this dissertation, we are interested in, from patients' perspective, the relative importance of 6 outcomes, so these 6 outcomes comprise our list of 6 objects. Often in each BWS task a set of these objects (e.g., 3 out of 6 outcomes), instead of all 6, is presented to individuals. Individuals are asked to complete repeated BWS tasks that are made up of different choice sets. Balance Incomplete Block Design is the design commonly used to generate the different choice sets. This design ensures that the occurrence and co-occurrence of objects across BWS tasks are constant. In our example, in each BWS task there are 3 outcomes for comparison, and individuals are asked to complete 10 BWS tasks.

BWS data can be analyzed using the "best minus worst score". For each object in a BWS study, we can calculate the number of times an object is chosen as "best" minus the number of times an object is chosen as "worst". This best minus worst score was found to have a linear relationship with the maximum likelihood estimates of the

[2] Flynn TN, Marley AAJ. Best-Worst Scaling: Theory and Methods. Handbook of Choice Modelling (in press)

conditional logit model, another approach to analyze BWS data.[3] One major advantage of using the best minus worst score for analysis is that it can help understand the choice data at individual level and thus facilitates exploring heterogeneity.[4]

[3] Marley AAJ, Louviere JJ. Some probabilistic models of best, worst, and best-worst choices. Journal of Mathematical Psychology. 2005;49:464-480.
[4] Auger P, Devinney TM, Louviere JJ. Using best-worst scaling methodology to investigate consumer ethical beliefs across countries. Journal of Business Ethics. 2007;70:299-326.

CHAPTER TWO

USE OF SURROGATE OUTCOMES IN FDA DRUG APPROVALS

Tsung Yu, Yea-Jen Hsu, Kevin M. Fain, Cynthia M. Boyd, Janet T. Holbrook, Milo A. Puhan

ABSTRACT

Objective

To evaluate how often surrogate outcomes are used as a basis for drug approvals by the US Food and Drug Administration (FDA) and whether and how the rationale for using treatment effects on surrogates as predictors of treatment effects on patient-centered outcomes was discussed.

Study design and Setting

We used the Drugs@FDA website to identify drug approvals produced from 2003 to 2012 by FDA. We focused on four diseases (chronic obstructive pulmonary disease (COPD), type 1 or 2 diabetes, glaucoma, and osteoporosis) where surrogates are commonly used in trials. We reviewed the drug labels and medical reviews.

Results

Of 1043 approvals screened, 58 (6%) were for the four diseases of interest. Most drugs for COPD (7/9, 78%), diabetes (25/26, 96%) and glaucoma (9/9, 100%) were approved based on surrogates while for osteoporosis, most drugs (10/14, 71%) were also approved for patient-centered outcomes (fractures). The rationale for using surrogates was discussed in 11 out of the 43 (26%) drug approvals based on surrogates.

Conclusion

Our results suggest that the FDA did not use a consistent approach to address surrogates in assessing benefits and harms of a drug. We proposed a framework for the use of surrogate outcomes in such assessments.

INTRODUCTION

The goal of treatments is to provide benefits to patients but at times treatments come with harms. For example, aspirin is prescribed for preventing cardiovascular diseases, but it is associated with an increased risk of gastrointestinal bleeds [1]. Thus, a comprehensive benefit-harm assessment is required to determine in which populations aspirin will have more benefits than harms [2, 3]. Benefit-harm assessment is also a key task for regulatory agencies because a new drug must be proven efficacious and safe for a specific indication before marketing [4]. The United States Food and Drug Administration (US FDA) and the European Medicines Agency (EMA) have both recognized the need to have a more structured and transparent approach to benefit-harm assessment that will inform the drug approval process for both agencies [5, 6]. Such an analysis begins by defining the decision-making context, including the appropriate population(s), intervention(s), comparison(s), outcome(s), and timeframe(s). Then, the drug reviewers conduct a quantitative analysis of treatment effects on different benefit and harm outcomes to make an overall assessment of the evidence. When doing benefit-harm assessment, selecting patient-centered outcomes, i.e., outcomes that patients notice and care about (survival, function, symptoms, and health-related quality of life) will make the results more informative to patients and other stakeholders [2, 7].

One major challenge of conducting a benefit-harm assessment is that surrogate outcomes are often used in randomized clinical trials (RCTs) to demonstrate the

treatment effects, without assessing patient-centered outcomes [8, 9]. A surrogate outcome is a biomarker or an intermediate to a patient-centered outcome that is "expected to predict clinical benefit (or harm or lack of benefit or harm)" [10]. The International Conference on Harmonisation issued guidelines for the conduct of clinical trials for the registration of drugs (ICH-9) that described a hierarchy of evidence for surrogacy [11]. In general, the evidence for surrogacy may come from pathophysiologic studies suggesting the biological plausibility of the association between surrogate outcomes and patient-centered outcomes, or from observational studies demonstrating the association between them. The highest level of evidence requires that RCTs have shown the treatment effects on surrogate outcomes can predict the treatment effects on patient-centered outcomes. The use of treatment effects on only surrogate outcomes for regulatory and clinical decision-making purposes remains commonplace regardless of the evidence supporting the validity of the surrogates [12, 13].

For prescription drugs treating certain diseases, surrogate outcomes were found to be used as primary outcome in about 50% of pivotal trials for regulatory approval [14]. For example, change in intraocular pressure (IOP) is often the primary outcome for RCTs on glaucoma medications instead of more direct measures of glaucoma such as optic disk deterioration or vision loss [15]. Another example is diabetes. Gandhi et al.[16] found that in 436 registered RCTs in type 1 or 2 diabetes, only 78 (18%) trials chose patient-centered outcomes as primary outcomes. Most trials used glycated hemoglobin to test the

efficacy of diabetes interventions instead of assessing their effects on clinical outcomes that have direct effects on patients' symptoms, function, and quality of life such as cardiovascular events. Some argue that the regulatory pathway of diabetes interventions relies too heavily on surrogates and should require patient-centered outcomes as well [17].

It is currently unclear if the FDA uses a consistent approach to address surrogate outcomes for drug approvals. A critical question is if the FDA review addresses whether treatment effects on surrogate outcomes predict the treatment effects on patient-centered outcomes across a spectrum of diseases. Our aim was to review the drug approvals for four diseases produced by the US FDA from 2003 to 2012 to learn how often these approvals were based on surrogate outcomes, and whether and how the rationale for using surrogate outcomes was discussed.

METHODS

Selection of drug approvals

We used the Drugs@FDA website (http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm) to identify all drug approvals produced from January 2003 to December 2012 (n=1043) by the US FDA. Drugs@ FDA is an open access database for drug products approved by the FDA that

contains a drug approval package, including information on drug labels, approval letters and reviews such as medical, chemistry, pharmacology, and statistical reviews. These reviews provide scientific analysis of a drug product and explain the FDA's thinking for the approval decision. Two authors (TY and YJH), working independently, screened the list to select the approvals that were eligible. The inclusion criteria were drug approvals where the drugs are indicated for the treatment of chronic obstructive pulmonary disease (COPD), diabetes, glaucoma, or osteoporosis. We focused on these four diseases because surrogate outcomes (lung function for COPD, blood sugar level for diabetes, IOP for glaucoma, and bone mineral density for osteoporosis respectively) are commonly used as primary outcomes in RCTs and all of them are "well established" surrogates for patient-centered outcomes [15, 18-20]. We excluded the drugs that are only indicated for a specific symptom related to the diseases or indicated for a specific patient population. Thus, we excluded a glaucoma drug that is indicated as an adjunct to ab externo glaucoma surgery, a diabetes drug approved for treating adult patients with endogenous Cushing's syndrome who have type 2 diabetes, and a drug treating diabetic peripheral neuropathic pain. We also removed any duplicate records. If there was a disagreement between the two authors about including or excluding a drug approval, we resolved it by discussion.

Data extraction

During the drug's approval process, the FDA review team will critically evaluate the drug's benefits and harms in different aspects and produce review documents, including the medical, chemistry, pharmacology, and statistical reviews, etc. For each included drug approval we retrieved the drug labels and medical reviews that were available on the Drug@FDA website. We focused on medical reviews instead of other reviews because medical reviews are, as we learned during pilot-testing of data extraction, most likely the review documents where the FDA reviewers address the issue of outcome selection. In addition, the medical review documents provide the FDA reviewers' assessment of clinical evidence that establishes the efficacy and safety of the drug. We developed and pilot-tested a standardized form for data extraction. Using the documents of drug labels and medical reviews, we extracted the information on indications and the primary outcomes that the indications were based on. We categorized these outcomes into a surrogate outcome or a patient-centered outcome, using the definition mentioned previously. For each drug approved based only on surrogates, we examined if the rationale for using surrogate outcomes was discussed or not (yes/no). We also assessed whether the surrogate was identified as being based on the highest level of evidence for surrogacy using the ICH-9 criteria, which is that the treatment effects on surrogates predict treatment effects on patient-centered outcomes. Two authors (TY and YJH) independently reviewed all documents and extracted the data. The discrepancies between authors on outcomes and the rationale for using surrogates outcomes were

resolved through discussion. Finally, we used descriptive statistics to summarize our findings.

RESULTS

Sixty-eight out of 1043 (7%) drug approvals were about COPD, diabetes, glaucoma or osteoporosis, and 58 (58/1043; 6%) of these were eligible for our study. The reasons for exclusion of approvals are summarized in **Figure 2-1**. Out of the 58 included approvals, 9 were for COPD (16%), 26 (45%) for diabetes, 9 (16%) for glaucoma, and 14 (24%) for osteoporosis. For three of the four examined conditions, the drug approvals were mostly based only on a surrogate outcome (COPD (7/9 approvals were based only on a surrogate, 78%), diabetes (25/26 approvals, 96%), and glaucoma (9/9 approvals, 100%) (**Table 2-1**). COPD drug approvals were primarily based on the effects on improving lung function, with the exception of two drug approvals (*SPIRIVA HANDIHALER* and *DALIRESP*) that also examined COPD exacerbations. Almost all diabetes drug approvals were based on lowering blood sugar level except for one drug approval (*JUVISYNC*), which included patient-centered outcomes (mortality and cardiovascular events) as well. Glaucoma drug approvals were all based on lowering IOP. Most drug approvals for osteoporosis (10/14; 71%) were based on both surrogate outcomes (bone mineral density) and patient-centered outcome (fractures).

Among the drugs that were approved based only on surrogates, 11 (11/43, 26%) of them discussed in the medical review the rationale for using surrogate outcomes to demonstrate drug efficacy for regulatory approval (**Table 2-2**). For COPD drug approvals based on surrogates, a medical review for one drug (*TUDORZA PRESSAIR*) mentioned the limitations of using lung function and the importance of evaluating patient-centered outcomes such as COPD exacerbations. For glaucoma, the reviews for three drugs (*ALPHAGAN P*, *QOLIANA*, and *LUMIGAN*) discussed the rationale for using change in IOP for drug approval. These reviews mentioned the association between high IOP and visual function loss but did not cite evidence from RCTs that an effect on IOP predicts an effect on visual function. For diabetes, we found that the reviews for seven drugs (*APIDRA, SYMLIN, EXUBERA, JANUVIA, JANUMET, VICTOZA,* and *BYDUREON*) discussed the rationale for use of surrogates and three of them (*SYMLIN, VICTOZA,* and *BYDUREON*) justified choosing glycemic control as an outcome by citing evidence from trials that corresponds to the highest level of evidence for surrogacy using the ICH-9 criteria. Specifically, they cited evidence that the treatment effect on blood glucose level can predict treatment effects on micro- or, to a lesser extent, macro-vascular complications. For example, in the review of *VICTOZA*, the reviewer stated that "*HbA1c has excellent reliability, predicts several diabetes-specific complications, and provides the current basis for treatment decisions. Lowering HbA1c reduces microvascular complications in patients with type 1 and type 2 diabetes and possibly macrovascular complications in patients with type 1 diabetes.*"

DISCUSSION

Our findings suggest that the FDA did not use a consistent approach to address surrogate outcomes when reviewing the drug approvals included in this study. Thus, the drug approvals are not consistently transparent about the role surrogate outcomes play in the drug approval process. For COPD, diabetes, and glaucoma, new drugs are often approved based on treatment effects on surrogate outcomes. But for osteoporosis, treatment effects on the surrogate outcome (bone mineral density) and the patient-centered outcome (fractures) are often examined together when regulatory decisions are made. The rationale for using surrogate outcomes for drug approval was not always discussed. If it was discussed, drug approvals for diabetes are more likely than drug approvals for the other examined conditions to contain a discussion of RCT evidence that demonstrates treatment effects on surrogate outcomes (blood sugar level) predict treatment effects on patient-centered outcomes (macro- or micro-vascular events).

Surrogate outcomes are sometimes chosen as the primary outcomes when designing the RCTs because they have the advantages of saving time, sample size, and resources in order to show a particular effect size compared to patient-centered outcomes that may take longer to accrue and require larger samples and longer studies [8]. In our survey, we found that in COPD, glaucoma, and diabetes, RCTs using surrogate outcomes also form the basis of drug approvals that were made by the FDA. The use of surrogate outcomes in RCTs, however, may complicate the benefit-harm assessment of treatments.

Surrogate outcomes are often captured using laboratory measurements, such as lung function for COPD, blood sugar level for diabetes and IOP for glaucoma. To calculate a benefit-harm metric like number needed to treat (NNT) or number needed to harm (NNH) [21], it is necessary to dichotomize the outcome variable to define a clinically meaningful change. For example, the target level for glycemic control is often set as hemoglobin A1C <7% in adult patients with type 2 diabetes mellitus, which is a level of hemoglobin A1C that has been linked to a lower risk of micro- or macro-vascular events [22]. However, using a threshold as an outcome is increasingly controversial because surrogates may have a continuous and non-linear relationship to the corresponding patient-centered outcomes [23]. Ideally, we would have treatment evidence on outcomes that are directly relevant to patients. RCTs should provide us direct evidence on how much a new diabetes drug lowers the risk of patient-centered outcomes such as stroke or amputation, and what side effects it causes to patients. Decision makers can then be better informed of whether or not treatment benefits outweigh the harms.

For the diseases we examined, FDA guidance documents provided guidelines for industry to design RCTs for regulatory approval [15, 18-20]. For COPD, diabetes, and glaucoma, the FDA allows using surrogate outcomes for drug approvals. For example, most glaucoma drugs are now indicated for "lowering IOP" but are not indicated for "treating glaucoma" (e.g., slowing glaucoma disease progression) [15]. Only the FDA guidance document for osteoporosis clarified that patient-centered outcomes (fractures)

are required for some classes of drugs to be approved for treating osteoporosis, because the association between increasing bone mineral density and lowering fracture risk has been shown to be inconsistent across trials [19, 24]. Hence, most approvals for osteoporosis drugs included in our study examined evidence on both bone mineral density and fractures.

We suggest that applications for almost all drugs should try to include direct evidence of the treatment effect on patient-centered outcomes. One may argue that it is not always practical to choose patient-centered outcomes as the primary outcome in RCTs because the events are too rare and the time period is too long to conduct the trial. It is thus unlikely that we will have data on patient-centered outcomes in every trial. If it is inevitable to use evidence on surrogate outcomes for drug applications, drug reviewers should at least review the level of the evidence for surrogacy (e.g., based on the ICH-9 criteria) when they are assessing the drugs' benefits and harms, and drug reviewers should also consider if the evidence for surrogacy is still relevant for the specific treatment and population under review.

For example, Staessen et al.[25] reviewed antihypertensive clinical trial data and examined the relationship between treatment effects on systolic blood pressure and treatment effects on cardiovascular events by drug class. Such a review can serve as the source for evidence for surrogacy when assessing treatment benefits and harms. In our

study samples, the number of drug approvals that provided the rationale for use of surrogate outcomes was lower than we expected and only some of them cited evidence from RCTs that evaluated the surrogates. It might be that the FDA concluded in prior drug approvals that the surrogate outcomes were valid and the rationale was provided (before 2003). At a minimum, FDA reviewers should cite those prior approvals. It would also be worthwhile for drug reviewers to address the evidence for a specific surrogate to be used for the drug and in the population under review. Further, quantifiable assumptions of surrogate effects on clinical outcomes will allow more valid and interpretable benefit-harm assessment of new drugs and also ease the communication of benefits and harms with patients and clinicians. Such an approach will provide greater transparency.

A proposal for a framework for use of surrogate outcomes in patient-centered benefit-harm assessment

Here we propose a framework (see **Figure 2-2**) for use of surrogate outcomes in patient-centered benefit-harm assessment:

This framework asks four questions. The 1st question in the framework is "*What are the corresponding patient-centered outcomes?*" It is important at first to be clear about which patient-centered outcomes the surrogate outcomes substitute for. If we could

use patient-centered outcomes in the benefit-harm assessment, what outcomes would be most relevant for the decision-making context? Then, the 2nd question in the framework is "*Are the corresponding patient-centered outcome data available?*" Data on surrogate outcomes and patient-centered outcomes can be both available to investigators. As we have described earlier for RCTs comparing treatments for osteoporosis, there are often data available on fracture events and bone mineral density. However, it has been shown that the association between increasing bone mineral density and lowering fracture risk is inconsistent [24]. Since data on facture events are available in trials, we can use this information and do not need to use data on bone mineral density. We suggest that if data on patient-centered outcomes are available, then they should be considered primary because they are the actual outcomes of interest and there seems no reason to use information from surrogate outcomes that may only partially capture the corresponding patient-centered outcomes.

If there are no patient-centered outcome data, we propose to ask the 3rd question *"What is the evidence supporting the use of surrogate outcomes as a substitute for patient-centered outcomes?"* To answer this question, we need to critically evaluate surrogate outcomes. The ICH-9 hierarchy of evidence can provide a framework for evaluation of surrogacy. Ideally, the evidence is based on RCTs that demonstrate the treatment effects on surrogate outcomes predict the treatment effects on patient-centered

outcomes. Moreover, one can conduct a systematic review, as Staessen et al.[25] did for antihypertensive drugs, of all available trial data to examine the association.

The 4th and the last question is "*Can the evidence for surrogacy be applied to the specific treatment and population under study?*" Even for the drugs that show a beneficial effect on surrogate outcomes, they are still likely to have unintended harmful effects on the patient-centered outcomes. For example, considering rosiglitazone, studies showed that although this drug effectively lowered the blood sugar level in patients with diabetes, it increased the risk of cardiovascular events [17]. When reviewing evidence of surrogacy, we should carefully consider if the evidence has been shown for the same drug (or the same class of drugs) and if the evidence can be applied to the specific treatment and population under review. It is possible that the evidence may no longer be appropriate as the scientific understanding evolves, or there are new issues or concerns about the use of the surrogate in certain populations.

These four questions will help us conduct patient-centered benefit-harm assessment when most of the efficacy data are about surrogate outcomes.

There are a few other published guidelines for evaluating surrogate outcomes in clinical trials. For example, Prentice proposed the statistical criteria for surrogate

outcomes that require them to "capture any relationship between the treatment and the true endpoint" [26]. Bucher et al. published a guidance document for surrogate outcomes as part of JAMA Users' Guides to the Medical Literature [27]. They proposed to ask several questions when addressing surrogate outcomes. First, evidence users should examine if the evidence supporting surrogate outcomes is valid; i.e., if there is a consistent association between treatment effect on the surrogate outcomes and treatment effect on the corresponding patient-centered outcomes. Second, evidence users should examine how "large, precise and lasting" of the treatment effect is on surrogate outcomes of interest. Finally, evidence users should assess if the treatment benefits caused by the change in surrogate outcomes outweigh the harms and costs to patients. Our framework agrees with what was proposed previously and we put more emphasis on the use of surrogate outcomes in the context of treatment benefit-harm assessment. We emphasize that patient-centered outcomes should be considered the primary data source in a benefit-harm assessment. When the use of surrogate outcomes is necessary, it is also important to review the evidence for surrogacy and consider if the evidence is still relevant for the specific treatment and population under review.

This study has limitations. We only focused on the surrogate outcomes used for drug efficacy and did not include surrogate outcomes that substitute for harms. Harmful events are often rare, so regulatory agencies may require evidence from studies beyond RCTs such as large and long-term post-marketing studies. We reviewed four diseases

where surrogate outcomes are commonly used but did not review diseases such as cancers or HIV where the use of surrogate outcomes is also prevalent. The validity of surrogate outcomes will vary by disease, and potentially by drug (or class of drugs) and by sub-population as well [8, 28, 29]. We did not evaluate the new drug applications that were declined by the FDA because these documents are not publicly available. There may be more explicit analysis of surrogate outcomes in those documents. We focused on medical reviews of the FDA drug approval process since we found that this is where a discussion of surrogate outcomes would most likely be documented but there is the possibility that it was mentioned somewhere else in the FDA reviews. Finally, not documenting the rationale for use of surrogate outcomes does not mean that the FDA reviewers did not take it into account when making decisions. However, a documented discussion of the evidence will certainly increase the transparency of the process in which regulatory bodies consider surrogate outcomes for drug approvals.

In conclusion, our findings suggest that for many chronic diseases, drugs are approved based on their treatment effects on surrogate outcomes, but that the FDA does not use a consistent approach for surrogates in order to evaluate these drug applications. This makes it difficult to assess and interpret their actual clinical effects relevant to patients. To conduct a patient-centered benefit-harm assessment, we should select patient-centered outcomes whenever possible. If the use of surrogate outcomes is

necessary, reviewing the evidence for surrogacy and considering its application in the treatment and population under study will substantially help us conduct such assessments.

REFERENCES

1. US Preventive Services Task Force. Aspirin for the prevention of cardiovascular disease: U.S. preventive services task force recommendation statement. Ann Intern Med. 2009 Mar 17;150(6):396-404.
2. Puhan MA, Singh S, Weiss CO, Varadhan R, Boyd CM. A framework for organizing and selecting quantitative approaches for benefit-harm assessment. BMC Med Res Methodol. 2012 Nov 19;12:173,2288-12-173.
3. Puhan MA, Singh S, Weiss CO, Varadhan R, Sharma R, Boyd CM. Evaluation of the Benefits and Harms of Aspirin for Primary Prevention of Cardiovascular Events: A Comparison of Quantitative Approaches [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013 Nov.
4. The Food and Drug Administration. Guidance for Industry. Providing Clinical Evidence of Effectiveness for Human Drugs and Biological Products. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm078749.pdf. 1998 (last accessed on February 1, 2014).
5. The European Medicines Agency. Development and testing of tools and processes for balancing multiple benefits and risks as an aid to informed regulatory decisions about medicinal products. Doc. Ref. EMEA/108979/2009. http://www.ema.europa.eu/docs/en_GB/document_library/Report/2011/07/WC500109477.pdf. 2009 (last accessed on February 1, 2014).
6. The Food and Drug Administration. Structured approach to benefit-risk assessment in drug regulatory decision-making. Draft PDUFA V implementation

plan - February 2013 fiscal years 2013-2017. http://www.fda.gov/downloads/ForIndustry/UserFees/PrescriptionDrugUserFee/UCM329758.pdf. 2013 (last accessed on February 1, 2014).

7. Methodology Committee of the Patient-Centered Outcomes Research Institute (PCORI). Methodological standards and patient-centeredness in comparative effectiveness research: the PCORI perspective. JAMA. 2012 Apr 18;307(15):1636-40.
8. Fleming TR, Powers JH. Biomarkers and surrogate endpoints in clinical trials. Stat Med. 2012 Nov 10;31(25):2973-84.
9. Svensson S, Menkes DB, Lexchin J. Surrogate outcomes in clinical trials: a cautionary tale. JAMA Intern Med. 2013 Apr 22;173(8):611-2.
10. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. Clin Pharmacol Ther. 2001 Mar;69(3):89-95.
11. International Conference on Harmonization. ICH harmonized tripartite guideline. Statistical principles for clinical trials - E9. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf. 1998 (last accessed on February 1, 2014).
12. Chakravarty A. Regulatory aspects in using surrogate markers in clinical trials. In: Burzykowsi T, Molenberghs G, Buyse M, eds. The evaluation of surrogate endpoints. Springer, 2005.

13. Mangiapane S, Velasco Garrido M. Use of surrogate end points in HTA. GMS Health Technol Assess. 2009;5:Doc12.
14. Downing NS, Aminawung JA, Shah ND, Krumholz HM, Ross JS. Clinical trial evidence supporting FDA approval of novel therapeutic agents, 2005-2012. JAMA. 2014 Jan 22-29;311(4):368-77.
15. Weinreb RN, Kaufman PL. The glaucoma research community and FDA look to the future: A report from the NEI/FDA CDER glaucoma clinical trial design and endpoints symposium. Invest Ophthalmol Vis Sci. 2009 Apr;50(4):1497-505.
16. Gandhi GY, Murad MH, Fujiyoshi A, Mullan RJ, Flynn DN, Elamin MB, et al. Patient-important outcomes in registered diabetes trials. JAMA. 2008 Jun 4;299(21):2543-9.
17. Rosen CJ. The rosiglitazone story--lessons from an FDA advisory committee meeting. N Engl J Med. 2007 Aug 30;357(9):844-6.
18. The Food and Drug Administration. Guidance for industry. Diabetes mellitus: developing drugs and therapeutic biologics for treatment and prevention. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM071624.pdf. 2008 (last accessed on February 1, 2014).
19. The Food and Drug Administration. Guidelines for preclinical and clinical evaluation of agents used in the prevention or treatment of postmenopausal osteoporosis. http://www.fda.gov/downloads/ScienceResearch/SpecialTopics/WomensHealthResearch/UCM131206.pdf. 1994 (last accessed on February 1, 2014).

20. The Food and Drug Administration. Guidance for industry. Chronic obstructive pulmonary disease: developing drugs for treatment. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071575.pdf. 2007 (last accessed on February 1, 2014).
21. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. N Engl J Med. 1988 Jun 30;318(26):1728-33.
22. American Diabetes Association. Standards of medical care in diabetes--2013. Diabetes Care. 2013 Jan;36 Suppl 1:S11-66.
23. Duckworth W, Abraira C, Moritz T, Reda D, Emanuele N, Reaven PD, et al. Glucose control and vascular complications in veterans with type 2 diabetes. N Engl J Med. 2009 Jan 8;360(2):129-39.
24. Seeman E. Is a change in bone mineral density a sensitive and specific surrogate of anti-fracture efficacy? Bone. 2007 Sep;41(3):308-17.
25. Staessen JA, Wang JG, Thijs L. Cardiovascular prevention and blood pressure reduction: A quantitative overview updated until 1 march 2003. J Hypertens. 2003 Jun;21(6):1055-76.
26. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. Stat Med. 1989 Apr;8(4):431-40.
27. Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA. Users' guides to the medical literature: XIX. applying clinical trial results. A. how to use an article measuring the effect of an intervention on surrogate end points. evidence-based medicine working group. JAMA. 1999 Aug 25;282(8):771-8.

28. Baker SG, Kramer BS. Surrogate endpoint analysis: an exercise in extrapolation. J Natl Cancer Inst. 2013 Mar 6;105(5):316-20.

29. Uhlig K, Leff B, Kent D, Dy S, Brunnhuber K, Burgers JS, et al. A Framework for Crafting Clinical Practice Guidelines that are Relevant to the Care and Management of People with Multimorbidity. J Gen Intern Med. 2014 Jan 18. [Epub ahead of print]

Figure 2-1. Review process

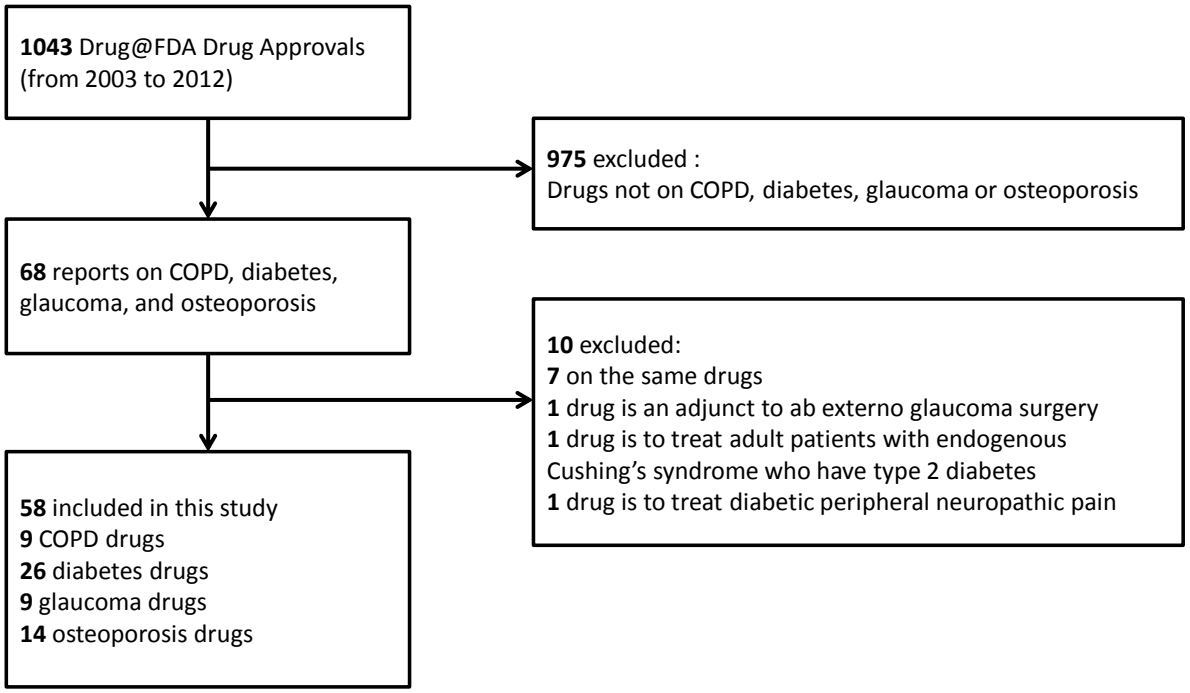


Figure 2-2. Framework for use of surrogate outcomes in patient-centered benefit-harm assessment

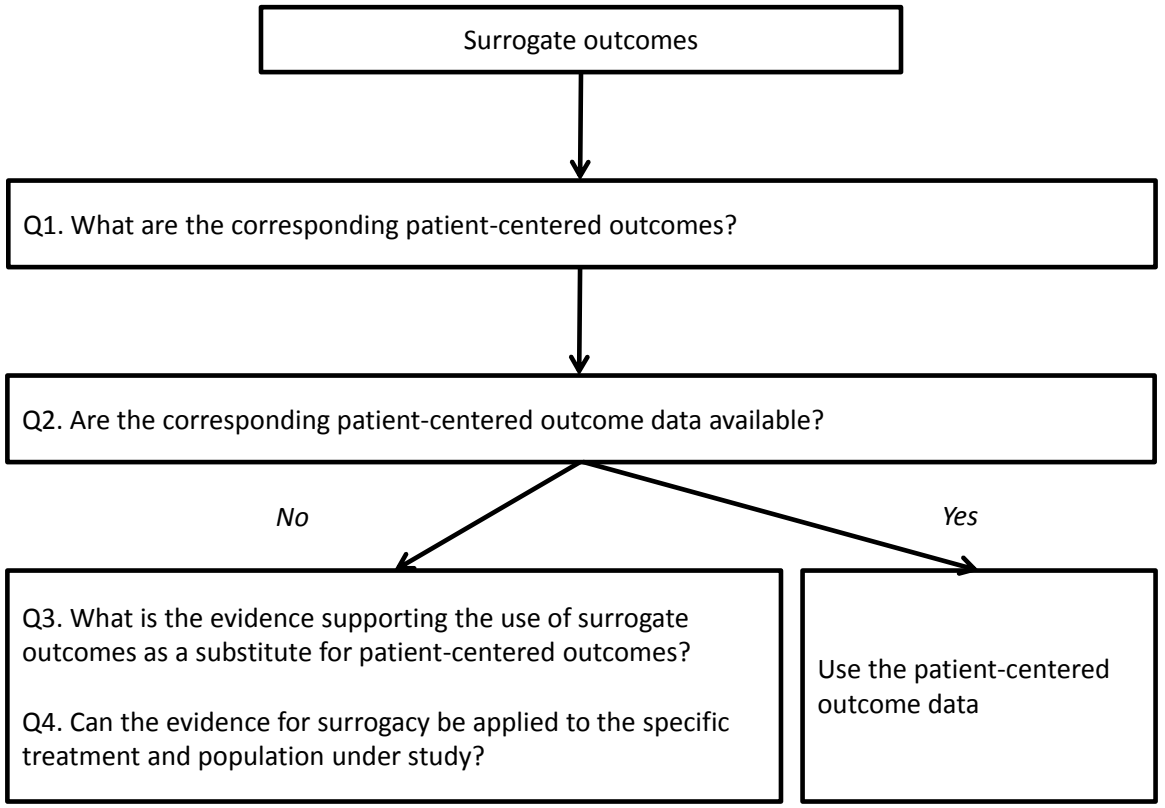


Table 2-1. Summary of the included drug approvals (n=58)

Disease	Drug name	Active ingredient	Year approved	Outcome that the indication is based on	Type of outcome that the indication is based on	Medical review discussed the rationale for using surrogate outcomes
COPD	ATROVENT HFA	Ipratropium bromide HFA	2004	Lung function	Surrogate outcome	No
COPD	SPIRIVA HANDIHALER	Tiotropium bromide inhalation powder	2004	Lung function COPD exacerbation	Both surrogate outcome and patient-centered outcome	Not applicable
COPD	BROVANA	Arformoterol tartrate	2006	Lung function	Surrogate outcome	No
COPD	SYMBICORT	Budesonide and formoterol fumarate dihydrate	2006	Lung function	Surrogate outcome	No
COPD	PERFOROMIST	Formoterol fumarate	2007	Lung function	Surrogate outcome	No
COPD	ARCAPTA NEOHALER	Indacaterol inhalation powder	2011	Lung function	Surrogate outcome	No
COPD	COMBIVENT RESPIMAT	Ipratropium bromide and albuterol	2011	Lung function	Surrogate outcome	No
COPD	DALIRESP	Roflumilast	2011	COPD exacerbation	Patient-centered outcome	Not applicable
COPD	TUDORZA PRESSAIR	Aclidinium bromide inhalation powder	2012	Lung function	Surrogate outcome	Yes
Diabetes (type 2)	RIOMET	Metformin hydrochloride oral solution	2003	Blood sugar level	Surrogate outcome	No
Diabetes (type 1 or 2)	APIDRA	Insulin glulisine [rDNA origin] injection	2004	Blood sugar level	Surrogate outcome	Yes
Diabetes (type 2)	FORTAMET	Metformin hydrochloride	2004	Blood sugar level	Surrogate outcome	No
Diabetes (type 2)	ACTOPLUS MET	Pioglitazone hydrochloride and metformin hydrochloride	2005	Blood sugar level	Surrogate outcome	No
Diabetes (type 2)	AVANDARYL	Rosiglitazone maleate and glimepiride	2005	Blood sugar level	Surrogate outcome	No
Diabetes	GLUMETZA	Metformin hydrochloride	2005	Blood sugar level	Surrogate outcome	No

Disease	Drug name	Active ingredient	Year approved	Outcome that the indication is based on	Type of outcome that the indication is based on	Medical review discussed the rationale for using surrogate outcomes
Diabetes (type 1 or 2)	LEVEMIR	Insulin detemir [rDNA origin] injection	2005	Blood sugar level	Surrogate outcome	No
Diabetes (type 1 or 2)	SYMLIN	Pramlintide acetate	2005	Blood sugar level	Surrogate outcome	Yes
Diabetes (type 2)	DUETACT	Pioglitazone hydrochloride and glimepiride	2006	Blood sugar level	Surrogate outcome	No
Diabetes (type 1 or 2)	EXUBERA	Insulin human [rDNA origin]	2006	Blood sugar level	Surrogate outcome	Yes
Diabetes (type 2)	JANUVIA	Sitagliptin	2006	Blood sugar level	Surrogate outcome	Yes
Diabetes (type 2)	JANUMET	Sitagliptin/metformin HCl	2007	Blood sugar level	Surrogate outcome	Yes
Diabetes (type 1 or 2)	NOVOLOG MIX 50/50	50% insulin aspart protamine suspension and 50% insulin aspart injection, [rDNA origin]	2008	Blood sugar level	Surrogate outcome	No
Diabetes (type 2)	PRANDIMET	Repaglinide and metformin HCl	2008	Blood sugar level	Surrogate outcome	No
Diabetes (type 2)	ACTOPLUS MET XR	Pioglitazone hydrochloride and metformin hydrochloride	2009	Blood sugar level	Surrogate outcome	No
Diabetes (type 2)	BYETTA	Exenatide	2009	Blood sugar level	Surrogate outcome	No
Diabetes (type 2)	CYCLOSET	Bromocriptine mesylate	2009	Blood sugar level	Surrogate outcome	No
Diabetes (type 2)	ONGLYZA	Saxagliptin	2009	Blood sugar level	Surrogate outcome	No
Diabetes (type 2)	WELCHOL	Colesevelam hydrochloride	2009	Blood sugar level Blood lipid level	Surrogate outcome	No
Diabetes (type 2)	KOMBIGLYZE XR	Saxagliptin and metformin HCl extended-release	2010	Blood sugar level	Surrogate outcome	No
Diabetes	VICTOZA	Liraglutide [rDNA origin]	2010	Blood sugar level	Surrogate outcome	Yes

Disease	Drug name	Active ingredient	Year approved	Outcome that the indication is based on	Type of outcome that the indication is based on	Medical review discussed the rationale for using surrogate outcomes
(type 2)		injection				
Diabetes (type 2)	JUVISYNC	Sitagliptin and simvastatin	2011	Blood sugar level Blood lipid level Mortality Cardiovascular disease	Both surrogate outcome and patient-centered outcome	Not applicable
Diabetes (type 2)	TRADJENTA	Linagliptin	2011	Blood sugar level	Surrogate outcome	No
Diabetes (type 2)	BYDUREON	Exenatide extended-release for injectable suspension	2012	Blood sugar level	Surrogate outcome	Yes
Diabetes (type 2)	JANUMET XR	Sitagliptin and metformin HCl extended-release	2012	Blood sugar level	Surrogate outcome	Medical review not available on the website
Diabetes (type 2)	JENTADUETO	Linagliptin and metformin hydrochloride	2012	Blood sugar level	Surrogate outcome	Medical review not available on the website
Glaucoma	ISTALOL	Timolol maleate ophthalmic solution	2004	Intraocular pressure	Surrogate outcome	No
Glaucoma	ALPHAGAN P	Brimonidine tartrate ophthalmic solution	2005	Intraocular pressure	Surrogate outcome	Yes
Glaucoma	QOLIANA	Brimonidine tartrate ophthalmic solution	2006	Intraocular pressure	Surrogate outcome	Yes
Glaucoma	TRAVATAN Z	Travoprost ophthalmic solution	2006	Intraocular pressure	Surrogate outcome	No
Glaucoma	COMBIGAN	Brimonidine tartrate/timolol maleate ophthalmic solution	2007	Intraocular pressure	Surrogate outcome	No
Glaucoma	ISOPTO CARPINE	Pilocarpine hydrochloride ophthalmic solution	2010	Intraocular pressure	Surrogate outcome	No
Glaucoma	LUMIGAN	Bimatoprost ophthalmic solution	2010	Intraocular pressure	Surrogate outcome	Yes
Glaucoma	COSOPT PF	Dorzolamide hydrochloride-timolol maleate ophthalmic solution	2012	Intraocular pressure	Surrogate outcome	No

Disease	Drug name	Active ingredient	Year approved	Outcome that the indication is based on	Type of outcome that the indication is based on	Medical review discussed the rationale for using surrogate outcomes
Glaucoma	ZIOPTAN	Tafluprost ophthalmic solution	2012	Intraocular pressure	Surrogate outcome	No
Osteoporosis (treatment and prevention)	BONIVA TABLETS	Ibandronate sodium	2003	Bone mineral density Fracture	Both surrogate outcome and patient-centered outcome	Not applicable
Osteoporosis (treatment and prevention)	FOSAMAX	Alendronate sodium	2003	Bone mineral density Fracture	Both surrogate outcome and patient-centered outcome	Not applicable
Osteoporosis (prevention)	PREMARIN	Conjugated estrogen tablets	2003	Bone mineral density	Surrogate outcome	No
Osteoporosis (prevention)	PREMPRO/PREMPHASE	Conjugated estrogens/medroxyprogesterone acetate tablets	2003	Bone mineral density	Surrogate outcome	No
Osteoporosis (prevention)	MENOSTAR	Estradiol transdermal system	2004	Bone mineral density	Surrogate outcome	No
Osteoporosis (treatment and prevention)	ACTONEL WITH CALCIUM	Risedronate sodium tablets with calcium carbonate tablets	2005	Bone mineral density Fracture	Both surrogate outcome and patient-centered outcome	Medical review not available on the website
Osteoporosis (treatment)	FORTICAL	Calcitonin-salmon [rDNA origin]	2005	Bone mineral density	Surrogate outcome	No
Osteoporosis (treatment)	FOSAMAX PLUS D	Alendronate sodium/cholecalciferol	2005	Bone mineral density Fracture	Both surrogate outcome and patient-centered outcome	Not applicable
Osteoporosis (treatment)	BONIVA INJECTION	Ibandronate sodium	2006	Bone mineral density Fracture	Both surrogate outcome and patient-centered outcome	Not applicable
Osteoporosis (treatment	EVISTA	Raloxifene hydrochloride	2007	Bone mineral density Fracture	Both surrogate outcome and patient-	Not applicable

Disease	Drug name	Active ingredient	Year approved	Outcome that the indication is based on	Type of outcome that the indication is based on	Medical review discussed the rationale for using surrogate outcomes
and prevention)					centered outcome	
Osteoporosis (treatment and prevention)	RECLAST	Zoledronic acid	2007	Bone mineral density Fracture	Both surrogate outcome and patient-centered outcome	Not applicable
Osteoporosis (treatment)	ATELVIA	Risedronate sodium	2010	Bone mineral density Fracture	Both surrogate outcome and patient-centered outcome	Not applicable
Osteoporosis (treatment)	PROLIA	Denosumab	2010	Bone mineral density Fracture	Both surrogate outcome and patient-centered outcome	Not applicable
Osteoporosis (treatment)	BINOSTO	Alendronate sodium	2012	Bone mineral density Fracture	Both surrogate outcome and patient-centered outcome	Medical review not available on the website

Table 2-2. Rationale for using surrogate outcomes discussed in drug medical reviews (n=11)

Disease	Drug name	Year approved	Rationale for using surrogate outcomes	The rationale is based on that "treatment effects on the surrogate outcome predict treatment effects on the patient-centered outcome" (highest level of evidence using ICH-9 criteria for surrogacy).
COPD	TUDORZA PRESSAIR	2012	*"Overall, the committee's view was that the Applicant's data for the primary endpoint of trough forced expiratory volume in one second (FEV1) demonstrated statistical significance, and that these results were clinically meaningful…Comments were made that the results for other measures of efficacy (e.g., the St. George's Respiratory Questionnaire [SGRQ] and COPD exacerbations), while generally not statistical significant, were nonetheless trending in a direction to support the results for the primary endpoint…Several comments were made regarding the limitations of FEV1-based endpoints and the importance of evaluating patient-centered outcomes."*	No

Disease	Drug name	Year approved	Rationale for using surrogate outcomes	The rationale is based on that "treatment effects on the surrogate outcome predict treatment effects on the patient-centered outcome" (highest level of evidence using ICH-9 criteria for surrogacy).
Diabetes (type 1 or 2)	APIDRA	2004	*"GHb (glycosylated hemoglobin) results were reported as glycated hemoglobin A1c (HbA1c) equivalents and are directly traceable to the Diabetes Control and Complications Trial (DCCT) reference, for which the relationship between mean BG (blood glucose) (measured by HbA1c) and the risk for vascular complications has been established."*	No
Diabetes (type 1 or 2)	SYMLIN	2005	*"That a reduction in HbA1c over 6-12 months can be associated with progression of diabetic retinopathy poses a regulatory dilemma. FDA (Food and Drug Administration) accepts reduction in HbA1c as a measure of efficacy in trials of new antidiabetic agents. This use of HbA1c as a surrogate endpoint reflects the finding that long-term reduction of HbA1c decreases the risk of diabetic complications, particularly retinopathy."*	Yes

Disease	Drug name	Year approved	Rationale for using surrogate outcomes	The rationale is based on that "treatment effects on the surrogate outcome predict treatment effects on the patient-centered outcome" (highest level of evidence using ICH-9 criteria for surrogacy).
Diabetes (type 1 or 2)	EXUBERA	2006	*"An ideal trial would use diabetic complications as endpoints, but the trial size and duration needed for use of such endpoints would be very large. There is some controversy about whether HbA1c is truly a good marker of the risk for complications of diabetes. However, the correlation of HbA1c with risk for the development of microvascular disease in Type 1 diabetics is well-established, and thus HbA1c is a good surrogate endpoint for the trials of inhaled insulin in Type 1 diabetics."*	No
Diabetes (type 2)	JANUVIA	2006	*"HbA1c is generally considered the most reliable surrogate of the glycemic control, and ultimately predicts late chronic complications of T2DM (type 2 diabetes mellitus) both microvascular and macrovascular, as demonstrated in the Diabetes Control and Complications Trial (DCCT) and the United Kingdom Prospective Diabetes*	Unclear

Disease	Drug name	Year approved	Rationale for using surrogate outcomes	The rationale is based on that "treatment effects on the surrogate outcome predict treatment effects on the patient-centered outcome" (highest level of evidence using ICH-9 criteria for surrogacy).
			Study (UKPDS)."	
Diabetes (type 2)	JANUMET	2007	*"HbA1c is generally considered the most reliable surrogate of the glycemic control, and ultimately predicts late chronic complications of T2DM (type 2 diabetes mellitus) both microvascular and macrovascular, as demonstrated in the Diabetes Control and Complications Trial (DCCT) and the United Kingdom Prospective Diabetes Study (UKPDS)."*	Unclear
Diabetes (type 2)	VICTOZA	2010	*"HbA1c has excellent reliability, predicts several diabetes-specific complications, and provides the current basis for treatment decisions. Lowering HbA1c reduces microvascular complications in patients with type 1 and type 2 diabetes and possibly macrovascular complications in patients*	Yes

Disease	Drug name	Year approved	Rationale for using surrogate outcomes	The rationale is based on that "treatment effects on the surrogate outcome predict treatment effects on the patient-centered outcome" (highest level of evidence using ICH-9 criteria for surrogacy).
			with type 1 diabetes."	
Diabetes (type 2)	BYDUREON	2012	*"HbA1c has excellent reliability, predicts several diabetes-specific complications, and provides the current basis for treatment decisions. Lowering HbA1c reduces microvascular complications in patients with type 1 and type 2 diabetes. There is weaker evidence showing that lowering HbA1c reduces macrovascular complications in patients with type 1 diabetes."*	Yes
Glaucoma	ALPHAGAN P	2005	*"Elevated IOP (intraocular pressure) presents a major risk factor in glaucomatous field loss. The higher the level of IOP, the greater the likelihood of optic nerve damage and visual field loss."*	No

Disease	Drug name	Year approved	Rationale for using surrogate outcomes	The rationale is based on that "treatment effects on the surrogate outcome predict treatment effects on the patient-centered outcome" (highest level of evidence using ICH-9 criteria for surrogacy).
Glaucoma	QOLIANA	2006	*"Elevated intraocular pressure is an etiological factor in glaucomatous cupping. Higher intraocular pressure corresponds with a greater frequency of optic nerve damage. Medical therapy for open-angle glaucoma is aimed at lowering the intraocular pressure below a level which is likely to produce further optic nerve damage."*	Unclear
Glaucoma	LUMIGAN	2010	*"Intraocular pressure (IOP) is currently the accepted standard for establishing the efficacy of ocular hypotensive medications. IOP is a surrogate end point for potential visual function loss."*	Unclear

CHAPTER THREE

PATIENTS WITH NON-INFECTIOUS UVEITIS CONSIDER IMPAIRED VISION, GLAUCOMA AND NEED FOR EYE SURGERY WORRYING ADVERSE TREATMENT OUTCOMES: RESULTS OF A BEST-WORST SCALING STUDY

ABSTRACT

Purpose

Local or systemic corticosteroid therapy in non-infectious uveitis is associated with different treatment outcomes. Little is known about how patients with non-infectious uveitis perceive the relative importance of these outcomes. We investigated and quantified patient preferences for outcomes using the best-worst scaling (BWS) approach.

Methods

We administered a preference-elicitation survey to patients with non-infectious uveitis who were in the Multicenter Uveitis Steroid Treatment Trial Follow-up Study (MUST FS) and to patients with a history of non-infectious uveitis who were treated in uveitis clinics at 2 academic medical centers (Johns Hopkins University and University of Pennsylvania). Participants were asked to complete 6 visual analogue scale (VAS) tasks as a warm-up exercise followed by 10 BWS tasks. For each BWS task they selected the most and the least worrying out of 3 outcomes. Outcomes included in the survey were vision not meeting the requirement for driving, diagnosed with cataracts, diagnosed with glaucoma, needing eye surgery, needing medicine for high blood pressure/cholesterol, and infection (e.g., sinusitis).

Results

Eighty-two patients in the MUST FS and 100 patients treated at the academic medical centers participated in the survey and provided valid responses for data analysis. According to BWS results, participants were more likely to select vision not meeting the requirement for driving (standardized individual BWS score: median = 80, interquartile range = 50 – 100), glaucoma (70, 60 – 90), and needing eye surgery (60, 50 – 80) as the most worrying treatment outcomes as compared against needing medicine for high blood pressure/cholesterol (30, 10 – 50), cataracts (30, 20 – 40), or infection (e.g., sinusitis) (20, 0 – 50). In regression analyses, race and education were identified as the main factors that influenced patient preferences.

Conclusions

Patients with non-infectious uveitis considered impaired vision, glaucoma and need for eye surgery worrying adverse treatment outcomes. Such preference data will provide important insights for making evidence- and preference- based treatment decisions.

INTRODUCTION

Results of the recent Multicenter Uveitis Steroid Treatment (MUST) Trial showed that both corticosteroid implant and systemic corticosteroids work to preserve vision for patients with non-infectious intermediate, posterior or panuveitis. But the MUST trial also indicated that each treatment is associated with different ocular and systemic adverse effects.[1, 2] A corticosteroid implant can be placed into a patient's eye to deliver corticosteroids locally for the management of patients with non-infectious uveitis.[3] These implants were designed to reduce the need for systemic use of corticosteroids, which increases the risk for, e.g., high blood pressure and infections. However, studies have shown that use of the implants leads to a higher risk of developing cataracts and glaucoma.[2-4] Based on these results it is difficult to choose between corticosteroid implant or systemic corticosteroids because both treatments have varying advantages and downsides.

Given that different treatments may lead to different outcomes, understanding how patients make trade-offs between these outcomes will substantially help inform benefit-harm assessments and treatment decisions.[5, 6] In a quantitative benefit-harm assessment, the aim is to compare treatments through a comprehensive assessment of treatment effects for multiple benefits and harms.[6] This method requires learning about patient preferences (i.e., the importance or weights that patients place on outcomes)[7] because the benefit-harm balance of treatments is sensitive to how patients value each

treatment outcome.[6, 8, 9] Besides benefit-harm assessment, preference data also play a key role in other health care research to inform policy, such as economic evaluation and decision analysis.[10, 11] Through incorporating patient preference data into these studies, research findings will reflect patients' needs and support patient-centered care.[12, 13]

The aim of our study was to learn how patients with non-infectious uveitis value treatment outcomes of corticosteroid implant and systemic corticosteroids by conducting a preference-elicitation survey. In addition, we explored whether patients' characteristics and their experiences with treatment outcomes influenced their preferences.

METHODS

Study Design

We conducted a cross-sectional survey and used best-worst scaling (BWS), a technique introduced by Finn and Louviere in marketing research,[14] to elicit patient preferences. We chose a "Case 1" BWS design, in which a set of 3 or more "objects" are shown to participants, who are asked to indicate their preferences by choosing, for example, the best and worst objects.[15, 16] Through studying the probability that participants choose an object over others, we can elicit the preferences that participants have for that object. BWS has increasingly been applied in health care research and it is

deemed less cognitively burdensome to participants than most other preference-elicitation methods.[17]

Patient Preferences Questionnaire

Through a literature review and consultation with investigators of the MUST Trial, we identified patient-centered treatment outcomes that are associated with corticosteroid implant and systemic corticosteroids regarding their use in non-infectious uveitis.[2, 18, 19] The main efficacy outcome was visual acuity and the main adverse outcomes included ocular events (cataracts, glaucoma, and eye surgery) and systemic events (hypertension, hyperlipidemia and systemic infections). We included 6 outcomes (see **Table 3-1**) in our preference-elicitation tasks and developed a description for each outcome with help from clinical experts and methodologists to make the wording as understandable as possible to patients. In our survey, we were interested in how patients on a group level make trade-offs between outcomes that are commonly seen in clinical practice. Some serious adverse events (e.g., endophthalmitis) are important to consider on an individual level when making treatment decisions but they are not as common so we decided to not include them in the list of outcomes.

The questionnaire started with asking participants to read the descriptions of the 6 outcomes and complete a visual analogue scale (VAS) task for each outcome. These VAS

tasks were designed to serve as a warm-up exercise in the questionnaire to introduce participants to the outcomes before they proceeded with the BWS tasks. After participants read the description of each outcome, they assigned a score from 0 (least serious outcome) to 100 (most serious outcome) on a VAS to indicate how serious they perceive the outcome to be.

Then the participants were asked to complete 10 BWS tasks. In each BWS, 3 outcomes ("objects" in BWS) were shown to participants and they were asked to choose the one that would worry them most and the one that would worry them least. To cover all 6 outcomes equally in the 10 BWS tasks, we used a balanced incomplete block design, as recommended by experts,[15] to generate 10 sets of outcomes (3 outcomes per set). The balanced incomplete block design assures each outcome would appear as an option equally often (5 times, $3 \times 10 \div 6$) in the questionnaire and also co-appear with each other equally often. Using data from BWS tasks, the preference for a treatment outcome relative to the others can be inferred statistically according to a series of choices that participants make. We did pilot-testing of the questionnaire with clinicians and students to ensure that the instructions were clear. The questionnaire that we administered can be found in the **Appendix 3-1**.

Study Participants

This study consisted of 2 populations of patients with non-infectious uveitis. For one population, 23 clinics in the Multicenter Uveitis Steroid Treatment Trial Follow-up Study (MUST FS) were contacted about their willingness to administer the Patient Preferences Questionnaire. Twelve clinics agreed to participate and administered the questionnaire to patients at their next scheduled follow-up visits between July 2013 and May 2014. The MUST Trial compared 2 active treatments, corticosteroid implant versus systemic corticosteroids, in patients with non-infectious intermediate, posterior and panuveitis. Details of its design and primary results can be found elsewhere.[1, 2] After the trial was concluded at 24 months after randomization, this cohort of patients has been followed for an additional 5 years for long-term treatment outcomes as part of the MUST FS. Institutional review boards (IRBs) at the coordinating center and at all 12 participating clinical centers provided their approvals for the addition of the Patient Preferences Questionnaire.

For the other population, outpatients with uveitis who were treated in ocular immunology clinics at the Sheie Eye Institute, University of Pennsylvania (referred to as PENN) or the Wilmer Eye Institute, Johns Hopkins University (referred to as JHU) were recruited. Between September 2013 and April 2014, patients at each clinic while waiting for their ophthalmologist appointment were contacted by study coordinators about their interest to complete the Patient Preferences Questionnaire. Inclusion criteria were

patients with a history of non-infectious anterior, intermediate, posterior or panuveitis and at least 18 years of age. Of note, we did not exclude patients with anterior uveitis although these patients are mainly treated with topical corticosteroids instead of corticosteroid implant or systemic corticosteroids. Because we believe that these patients also can contribute valid data on patient preferences for treatment outcomes. If patients agreed to participate and provided their oral consent, we administered the questionnaire to them and, additionally, asked 14 respondent-specific questions on their demographic information (e.g., age, gender, race, and education) and clinical characteristics (e.g., systemic comorbidities, vision, and experiences with treatments and outcomes). All study procedures were approved by IRBs at both institutions.

Data Analysis

To analyze BWS data, we assigned a "+1" to the outcome chosen as the most worrying one and a "-1" to the outcome chosen as the least worrying one in each scenario. We computed an "individual BWS score" for each of the 6 outcomes (number of times an outcome was picked as the most worrying by a participant minus the number of times it was picked as the least worrying). In the questionnaire, each outcome appeared 5 times across 10 BWS tasks. Thus, the individual BWS score for each outcome was on a scale bounded by -5 and +5, and the larger the score, the more worrying the outcome to the individual. Additionally, we counted the occurrence of best and worst choices across all participants to calculate the "aggregate BWS score" for each outcome. We chose to

analyze our data using BWS scores because BWS scores are easy to interpret and they have been shown to provide sufficient statistics for various regression models.[15] We were interested in learning if and how preference data elicited using the BWS method were different from that using the VAS method. Hence, to facilitate the comparison, we standardized the individual BWS scores on a 0-100 scale (same as the VAS scores). We constructed box plots of the standardized individual BWS scores and the VAS scores and ranked the outcomes based on the median of the scores.

We used simple linear regression to compare the 6 standardized individual BWS scores between the 2 populations (MUST FS vs PENN or JHU) and to explore the associations between each of the 6 BWS scores and patient characteristics that may influence their preferences (gender, age, race, education, time since diagnosis, location of uveitis, and experiences with treatment outcomes). To account for potential confounding, we constructed 6 multiple linear regression models with BWS scores of each outcome as the dependent variables, and in each model we adjusted for patient characteristics simultaneously. For each patient characteristic variable, we recognized that we were examining multiple associations between the variable and each of the 6 dependent variables. Thus, we applied Bonferroni correction to the statistical significance level (a p-value of 0.05 divided by 6, 0.008) to minimize the problem of multiplicity when doing tests for significance. Analyses were conducted using SAS 9.3 (SAS Inc, Cary, NC) and Stata 11.2 (StataCorp LP, Colllege Station, TX).

RESULTS

Patient Characteristics

As of July 2013, 210 patients were under follow-up in the MUST FS. Between July 2013 and May 2014, 89 patients across 12 MUST FS clinical centers completed the Patient Preference Questionnaire at their study visits. Seven patients did not provide valid responses on BWS tasks (in which the most and least worrying outcomes were not clearly indicated), and they were omitted from our analysis. Of the 82 patients included in our analysis, 36 were originally assigned to the implant therapy group and 46 to the systemic therapy group. Between September 2013 and April 2014, another 107 outpatients with non-infectious uveitis (73 from PENN and 34 from JHU) were also recruited for this study and 100 patients (68 from PENN and 32 from JHU) completed the questionnaire and provided valid responses. Taken together, our final analysis was based on responses of 182 patients with non-infectious uveitis.

The socio-demographic, disease, and treatment-related characteristics of the 182 study participants are shown in **Table 3-2**. The majority of these 182 participants were female (74%), white (58%), employed with income (57%), and with high school graduate (30%) or some college education (29%). The mean age was 52 years (standard deviation 15). The distributions of age and employment status were not significantly different between patients recruited from the MUST FS and patients recruited from PENN or JHU, but there were more male (15% vs 35%) and black (29% vs 44%) participants and more

people having high educational level (college graduate or higher: 24% vs 47%) in the latter population. The MUST trial did not include patients with anterior uveitis, while most patients we recruited from PENN or JHU were patients with a history of anterior uveitis (52%), who were less severe cases. Patients from the MUST FS were more likely to be bilateral (88% vs 72%), to have been diagnosed with uveitis for a longer time (median years: 10 vs 5), and to have more experiences with treatments including corticosteroid injections, systemic corticosteroids, and eye surgery. But their experiences with certain diseases (associated systemic diseases, high blood pressure, high cholesterol, and glaucoma) was not significantly different from patients recruited from PENN or JHU, except for cataracts (more patients had cataracts in MUST FS). Among the 182 study participants, 23% had been diagnosed with associated systemic disease; 42% had high blood pressure; 36% had high cholesterol; 31% had glaucoma and 80% had cataracts. Their vision status can also be found in **Table 3-2**.

Preferences Scores

Estimates of treatment outcome importance using BWS are presented in **Table 3-3**. We counted the total number of times across all surveys that each object (treatment outcome) was chosen as most worrying or least worrying, or was not chosen and computed the aggregate BWS score for each outcome. The number of counts was divided by the availability of each object (N=910, the number of times that an object appeared across 10 BWS tasks and 182 participants) to calculate a proportion. The aggregate BWS

score for glaucoma was the largest (387) and the score for cataracts was the smallest (-333). Individual BWS scores for each outcome were calculated and the median and interquartile range of the scores (bounded by -5 and +5) and the standardized scores (0-100) are also presented.

Box plots of the standardized individual BWS scores and the VAS scores (on a scale from 0 to100) are shown in the **Figure 3-1**. According to the standardized individual BWS scores of the 6 outcomes, we identified that one group of outcomes, including vision not meeting the requirement for driving (median: 80), glaucoma (median: 70), and needing eye surgery (median: 60), were considered more worrying by study participants than the other group, including needing medicine for high blood pressure/cholesterol (median: 30), cataracts (median: 30), and infection (median: 20). Distributions of the scores of vision not meeting the requirement for driving and infection (e.g., sinusitis) had the largest variability as shown by interquartile ranges.

With regard to the VAS scores, the median score of each outcome was 80 (vision not meeting the requirement for driving), 85 (glaucoma), 70 (needing eye surgery), 70 (needing medicine for high blood pressure/cholesterol), 70 (cataracts), and 60 (infection) respectively. The ranking of the 6 outcomes by median of the standardized individual BWS scores was different from the ranking by median of the VAS scores. Comparing the distributions of medians of these 2 types of scores (see **Figure 3-1**), it suggests that using

BWS methods to elicit patient preferences seems easier to differentiate the importance of outcomes than using VAS methods.

Associations between Patient Characteristics and Preferences Scores

Table 3-4 shows the results from bivariate analyses that examined the associations between each patient characteristic and standardized individual BWS scores of the 6 outcomes. Bivariate analyses show that there were no significant differences in each score between the 2 populations (MUST FS vs PENN or JHU). For socio-demographic variables, gender and age were not significantly associated with the scores of any outcome. Race (regression coefficient: 21.0; 95% CI: 11.6 – 30.4) and education (regression coefficient: 14.5; 95% CI: 4.6 – 24.5) were significantly (using 0.008 as statistical significance level) associated with BWS score of vision not meeting requirement for driving, and education (regression coefficient: -16.9; 95% CI: -26.4 – -7.3) was significantly associated with the score of infection (e.g., sinusitis). Disease characteristics such as time since diagnosis and location of uveitis (anterior vs other) were not significantly associated with BWS scores of any outcome. Furthermore, when we compared the BWS scores between patients who had the respective outcome previously (cataracts, glaucoma, eye surgery, or high blood pressure/cholesterol) versus patients who did not, none of the associations reached the 0.008 level of significance. **Table 3-5** shows the results of the 6 multiple linear regression models with each standardized individual BWS score as the dependent variables and in which we adjusted

for all characteristics simultaneously. The results were similar to those from the bivariate analyses and suggested that only race and education may influence patient preferences for treatment outcomes.

DISCUSSION

We conducted a preference-elicitation survey of patients with non-infectious uveitis to measure the relative importance of treatment outcomes. Findings from BWS suggested that patients considered treatment outcomes including vision not meeting the requirement for driving, glaucoma, and needing eye surgery more worrying as compared against cataracts, needing medicine for controlling high blood pressure/cholesterol, or infection (e.g., sinusitis).

Our study provides useful information to decision makers, including patients, clinicians, and policy makers, about how patients themselves perceive the relative importance of outcomes that are crucial to making treatment decisions in non-infectious uveitis. Participants in our study generally considered ocular adverse events more worrying than systemic adverse events, except that cataracts were viewed the least worrying. Interestingly, based on the aggregate BWS scores, glaucoma was considered the most worrying by participants. Glaucoma and cataracts are both the common complications of uveitis and also side effects from corticosteroid therapy.[20] But

perhaps influenced by different disease prognoses and different abilities of the medical community to manage these diseases,[21, 22] our study participants had substantially different preferences for them.

One important methodological issue in a preference-elicitation study is the question about in which patient population the preferences should be elicited and if patients' prior experiences with treatments or outcomes affect the ratings.[23, 24] When designing the study, we chose to define broad inclusion criteria and enrolled 2 populations in order to have the opportunity to explore patient characteristics. One population (MUST FS) included more severe cases of uveitis and those patients had long-term experiences with treatments and outcomes. In the other population (PENN or JHU), most patients had less severe disease (anterior uveitis) and were less experienced with treatments or outcomes. We found, however, there were no significant differences between the 2 patient populations (MUST FS vs PENN or JHU). According to our regression analyses, the factors that influence preferences the most were socio-demographic factors. For example, white patients in our study had significantly higher preference scores for vision not meeting the requirement for driving. Patients with higher educational levels had significantly higher preference scores for vision not meeting the requirement for driving but significantly lower preference scores for infection. For other variables examined, we did not detect any significant association between patients' disease characteristics or prior experiences with outcomes and their preference scores.

The results remained similar even when we adjusted for all patient characteristic variables at the same time.

One of the challenges we encountered while designing the questionnaire was to ensure that participants had a common understanding of the treatment outcomes when they were doing preference-elicitation tasks. To achieve this goal, we asked participants to read the description of each outcome and complete the VAS tasks as a warm-up exercise before they completed the BWS tasks. Still, the variability of preferences between individuals was large. By examining the distribution of individual BWS scores, we found the interquartile ranges for the outcomes vision not meeting the requirement for driving and infection (e.g., sinusitis) were the largest. In the questionnaire, we associated vision with driving standard since vision is a rather abstract concept to describe. We found in our study white patients and patients with higher educational levels considered vision not meeting the requirement for driving more worrying than other patients. But we should be cautious about this finding since the association may be confounded by if and how frequent the participants drive, which we did not assess. We restricted the description of infection to sinusitis (most common one) since different kinds of infection can be perceived differently by patients. We found the preferences for infection still varied appreciably among participants. We were unsure if this reflected the actual distribution of the scores or if participants were thinking about infection in different ways.

As for comparing the results from VAS with BWS, we found the difference in the median of VAS scores for each outcome (ranged from 60 to 85) was smaller than the difference in the median of standardized individual BWS scores (20 to 80). One advantage of using BWS to elicit patient preferences is that it allowed us to ask participants in a way that they can make trade-offs between objects.[15] In contrast, in doing VAS tasks, there are no trade-offs involved so that method may be less sensitive to detect differences in the scores of the objects being rated.[25] In addition, the anchors (e.g., most or least serious outcome in our case) of VAS may have various meanings to individuals, which may affect our findings. We have demonstrated in this study that most participants can complete BWS tasks to indicate their preferences without major difficulty. Investigators who plan to elicit patient preferences may consider using this approach for their future studies.

Preference data such as what we collected are essential to benefit-harm assessment. As in our example, corticosteroid implant and systemic corticosteroids for non-infectious uveitis are associated with distinct systemic and ocular adverse events, so it is difficult to make decisions without doing a comprehensive benefit-harm assessment. Doing a quantitative benefit-harm assessment where multiple outcomes are involved inevitably requires assigning the relative importance to each outcome. With patient preference data, we are better informed of how patients, the most important stakeholder, make trade-offs between treatment benefits and harms. There is a growing interest in both

US Food and Drug Administration and European Medicines Agency in incorporating patients' perspectives into their regulatory process, such as review of prescription drugs for marketing.[26] The current practice is somewhat limited to consulting patient groups or including patients on advisory committees but rarely uses such information quantitatively and explicitly.[13] Our study demonstrated an alternative and more explicit approach to incorporating the patients' perspectives. We surveyed the key patient population who are most familiar with the condition and elicited their preferences for outcomes quantitatively. These data can be combined with clinical trial data to estimate the benefit-harm balance of competing treatments and provide an effective way to engage patients in the process of developing evidence that informs preference-based decisions.

A number of limitations were identified in this study. We elicited the "stated preferences" from patients, in which patients completed preference-elicitation tasks based on their judgment of hypothetical descriptions of outcomes. Communicating with patients about treatment outcomes is a challenging task because the descriptions should include enough information to describe the outcome while not being too complicated for patients to understand. An individual participant's familiarity with outcomes also may affect how these outcomes are perceived,[5] which is difficult to measure. In our study sample we included patients with anterior uveitis for whom either systemic or implant therapy may not be indicated; however, we did not observe significantly different preferences in patients with anterior uveitis from other patients. This, in fact, also assured us that our

findings were consistent across patient groups to some extent and the participants were not making choices randomly on BWS tasks. We hypothesized that patients' experiences with outcomes may play a role in their preferences, but with our study sample size, we had limited power for testing for these hypotheses. Our study aim was to elicit preferences for patient-centered outcomes that are commonly seen in practice and measured in clinical studies. Thus, we developed our questionnaire using review of the existing literature and consultation with clinical experts. Another, probably better way to develop the questionnaire would be to conduct qualitative research with patients (e.g., focus group[27]) at the beginning. This would ensure that outcomes that are meaningful to patients are captured in the preference-elicitation tasks.

In summary, we have demonstrated a feasible approach to eliciting patient preferences for treatment outcomes. Participants in our study were comfortable and willing to complete BWS to indicate their preferences for outcomes. Patients with non-infectious uveitis considered vision not meeting the requirement for driving, glaucoma, and needing eye surgery more worrying treatment outcomes as compared against cataracts, needing medicine for high blood pressure/cholesterol, or infection (e.g., sinusitis). Data on these patient preferences will provide important insights for patients and clinicians to make evidence- and preference-based treatment decisions.

REFERENCES

1. Multicenter Uveitis Steroid Treatment Trial Research Group, Kempen JH, Altaweel MM, Holbrook JT, Jabs DA, Sugar EA. The multicenter uveitis steroid treatment trial: rationale, design, and baseline characteristics. Am J Ophthalmol. 2010;149(4):550-561.e10.
2. Multicenter Uveitis Steroid Treatment (MUST) Trial Research Group, Kempen JH, Altaweel MM, Holbrook JT, Jabs DA, Louis TA, Sugar EA, Thorne JE. Randomized comparison of systemic anti-inflammatory therapy versus fluocinolone acetonide implant for intermediate, posterior, and panuveitis: the multicenter uveitis steroid treatment trial. Ophthalmology. 2011;118(10):1916-26.
3. Sallam A, Taylor SR, Lightman S. Review and update of intraocular therapy in noninfectious uveitis. Curr Opin Ophthalmol. 2011;22(6):517-22.
4. Friedman DS, Holbrook JT, Ansari H, Alexander J, Burke A, Reed SB, Katz J, Thorne JE, Lightman SL, Kempen JH; MUST Research Group. Risk of elevated intraocular pressure and glaucoma in patients with uveitis: results of the multicenter uveitis steroid treatment trial. Ophthalmology. 2013;120(8):1571-9.
5. Brett Hauber A, Fairchild AO, Reed Johnson F. Quantifying benefit-risk preferences for medical interventions: an overview of a growing empirical literature. Appl Health Econ Health Policy. 2013 Aug;11(4):319-29.
6. Puhan MA, Singh S, Weiss CO, Varadhan R, Boyd CM. A framework for organizing and selecting quantitative approaches for benefit-harm assessment. BMC Med Res Methodol. 2012;12:173,2288-12-173.

7. Joy SM, Little E, Maruthur NM, Purnell TS, Bridges JF. Patient preferences for the treatment of type 2 diabetes: a scoping review. Pharmacoeconomics. 2013;31(10):877-92.
8. Yu T, Vollenweider D, Varadhan R, Li T, Boyd C, Puhan MA. Support of personalized medicine through risk-stratified treatment recommendations - an environmental scan of clinical practice guidelines. BMC Med. 2013 Jan 9;11:7.
9. Yu T, Fain K, Boyd CM, Singh S, Weiss CO, Li T, Varadhan R, Puhan MA. Benefits and harms of roflumilast in moderate to severe COPD. Thorax. 2013 Dec 17 [Epub ahead of print].
10. Neumann PJ, Goldie SJ, Weinstein MC. Preference-based measures in economic evaluation in health care. Annu Rev Public Health. 2000;21:587-611.
11. Sommers BD, Beard CJ, D'Amico AV, Dahl D, Kaplan I, Richie JP, Zeckhauser RJ.Cancer. Decision analysis using individual patient preferences to determine optimal treatment for localized prostate cancer. 2007 Nov 15;110(10):2210-7.
12. Montori VM, Brito JP, Murad MH. The optimal practice of evidence-based medicine: incorporating patient preferences in practice guidelines. JAMA. 2013;310(23):2503-4.
13. van Til JA, Ijzerman MJ. Why should regulators consider using patient preferences in benefit-risk assessment? Pharmacoeconomics. 2014 Jan;32(1):1-4.
14. Finn A and Louviere JJ. Determining the Appropriate Response to Evidence of Public Concern: The Case of Food Safety. J Public Pol Mark. 1992;11(1):12-25.

15. Louviere JJ, Lings I, Islam T, Gudergan S, Flynn T. An introduction to the application of (case 1) best–worst scaling in marketing research. Int J Res Mark. 2013;30(3):292–303.
16. Louviere JJ, Flynn TN. *Using best-worst scaling choice experiments to measure public perceptions and preferences for healthcare reform in Australia.* Patient. 2010 Dec 1;3(4):275-83.
17. Flynn TN. Valuing citizen and patient preferences in health: recent developments in three types of best-worst scaling. Expert Rev Pharmacoecon Outcomes Res. 2010 Jun;10(3):259-67.
18. Da Silva JA, Jacobs JW, Kirwan JR, Boers M, Saag KG, Inês LB, de Koning EJ, Buttgereit F, Cutolo M, Capell H, Rau R, Bijlsma JW. Safety of low dose glucocorticoid treatment in rheumatoid arthritis: published evidence and prospective trial data. Ann Rheum Dis. 2006;65(3):285-93.
19. Sarnes E, Crofford L, Watson M, Dennis G, Kan H, Bass D. Incidence and US costs of corticosteroid-associated adverse events: a systematic literature review. Clin Ther. 2011 Oct;33(10):1413-32.
20. Carnahan MC, Goldstein DA. Ocular complications of topical, peri-ocular, and systemic corticosteroids. Curr Opin Ophthalmol. 2000 Dec;11(6):478-83.
21. American Academy of Ophthalmology Cataract and Anterior Segment Panel. Preferred Practice Pattern Guidelines. Cataract in the Adult Eye. San Francisco, CA: American Academy of Ophthalmology; 2011. Available at: www.aao.org/ppp.

22. American Academy of Ophthalmology Glaucoma Panel. Preferred Practice Pattern Guidelines. Primary Open-Angle Glaucoma. San Francisco, CA: American Academy of Ophthalmology; 2011. Available at: www.aao.org/ppp.
23. De Wit GA, Busschbach JJ, De Charro FT. Sensitivity and perspective in the valuation of health status: whose values count? Health Econ. 2000 Mar;9(2):109-26.
24. Ubel PA, Loewenstein G, Jepson C. Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. Qual Life Res. 2003 Sep;12(6):599-607.
25. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the economic evaluation of health care programmes. Oxford university press, 2005.
26. The Food and Drug Administration. Structured approach to benefit-risk assessment in drug regulatory decision-making. Draft PDUFA V implementation plan - February 2013 fiscal years 2013-2017. http://www.fda.gov/downloads/ForIndustry/UserFees/PrescriptionDrugUserFee/UCM329758.pdf. 2013 (last accessed on April 23, 2014).
27. von Wagner C1, Halligan S, Atkin WS, Lilford RJ, Morton D, Wardle J. Choosing between CT colonography and colonoscopy in the diagnostic context: a qualitative study of influences on patient preferences. Health Expect. 2009 Mar;12(1):18-26.

Table 3-1. Objects for the best-worst scaling tasks

Treatment outcomes ("objects" in best-worst scaling)	Description to patients
Vision not meeting the requirement for driving	20/20 is a way to express normal vision. This means that at 20 feet you can see clearly what should be seen at that distance. To obtain a driver's license, your vision with glasses correction must be at least 20/40 (meaning that the letters you can read at 20 feet away can be read correctly by people with normal vision at 40 feet away). You cannot drive on public roads if your vision does not meet this requirement.
Cataracts	A cataract is a clouding of the lens in the eye. Patients may not notice that they have cataracts at first. As the eye lens becomes more and more clouded, vision becomes impaired and may result in blurry vision. Other symptoms of cataracts include having poor night vision or seeing halos around lights. Cataract surgery may be recommended by your doctor. It is usually safe and successful in restoring vision.
Glaucoma	Glaucoma is an eye disease that can lead to vision loss. Over time patients with glaucoma can lose side vision and have reduced ability to see shades of gray. Moderate glaucoma can lead to reduced mobility, slower and more difficult reading. Treatments include medications, like daily eye drops, and sometimes surgery. There is no cure for glaucoma. Once someone has it they will probably have to take medications for it for the rest of their life unless they have surgery.
Needing eye surgery (for cataracts or high eye pressure)	Patients sometimes develop problems with their eyes that need surgery, for example, a cataract. Cataract surgery is done with a painless operation that takes about 30 minutes. It does not require you to stay overnight in the hospital. Other patients may need surgery to lower the pressure in their eyes. This type of surgery takes about 45 minutes. You can go home the same day just as with cataract surgery. Both surgeries usually work well, but any eye surgery has risks including eye infection which can rarely lead to vision loss.
Needing medicine for controlling high blood pressure or cholesterol	High blood pressure and high cholesterol (fats in the blood) can lead to heart attack or stroke. Keeping your blood pressure and cholesterol under control can reduce your risk of heart disease. Many medicines can help you control blood pressure or cholesterol. For example, if you have high blood pressure, doctors may ask you to take antihypertensive drugs. If you have high cholesterol, they may ask you to take a statin drug. There can be side effects of these drugs such as headache and fatigue but they are generally mild.
Infection (e.g., sinusitis)	Some drugs work by suppressing your immune system, so after taking the drugs you

	may be at a higher risk of infections. For example, sinusitis is one type of infection commonly seen. When you have sinusitis, your nose is stuffy. You may cough, feel tired, get a headache or have a fever. Doctors may ask you to take antibiotics to control the infection. Usually infections resolve quickly after they are treated.

Table 3-2. Socio-demographic, disease and treatment-related characteristics of the study participants

Characteristics	Total	Patient population		P-value*
		MUST FS	PENN or JHU	
Number of participants	182	82	100	
Male gender, n (%)	47 (26%)	12 (15%)	35 (35%)	0.002
Age, mean (SD)	52 (15)	53 (14)	51 (16)	0.637
Race, n (%)				
White	105 (58%)	52 (63%)	53 (53%)	0.077
Black	68 (37%)	24 (29%)	44 (44%)	
Other	9 (5%)	6 (7%)	3 (3%)	
Hispanic, n (%)	9 (5%)	8 (10%)	1 (1%)	0.012
Employment status, n (%)				
Employed with income	103 (57%)	47 (57%)	56 (56%)	0.451
Homemaker	7 (4%)	5 (6%)	2 (2%)	
Student	6 (3%)	2 (2%)	4 (4%)	
Retired	33 (18%)	11 (13%)	22 (22%)	
Disabled	24 (13%)	12 (15%)	12 (12%)	
Unemployed	9 (5%)	5 (6%)	4 (4%)	
Education, n (%)				
Grade 11 or less	10 (5%)	7 (9%)	3 (3%)	0.010
Grade 12 or high school graduate	54 (30%)	28 (34%)	26 (26%)	
Some college	52 (29%)	28 (34%)	24 (24%)	
College graduate	35 (19%)	12 (15%)	23 (23%)	
One or more years post college	31 (17%)	7 (9%)	24 (24%)	
Location of uveitis, n (%)				
Anterior uveitis	52 (29%)	-	52 (52%)	N/A
Anterior and intermediate uveitis	6 (3%)	-	6 (6%)	
Intermediate uveitis	42 (23%)	25 (30%)	17 (17%)	
Posterior or panuveitis	82 (45%)	57 (70%)	25 (25%)	
Bilateral uveitis, n (%)	144 (79%)	72 (88%)	72 (72%)	0.009
Time since diagnosis in years, median (IQR)	8 (5-13)	10 (8-14)	5 (2-10)	<0.001
Treatment experience, n (%)				
Periocular or intraocular injections for uveitis	82 (45%)	49 (60%)	33 (33%)	<0.001
Systemic therapy for uveitis	124 (68%)	81 (99%)	43 (43%)	<0.001
Eye surgery	136 (75%)	78 (95%)	58 (58%)	<0.001
Disease experience, n (%)				
Associated systemic disease	41 (23%)	18 (22%)	23 (23%)	0.866
High blood pressure	77 (42%)	34 (41%)	43 (43%)	0.835
High cholesterol	65 (36%)	32 (39%)	33 (33%)	0.399

Glaucoma	57 (31%)	29 (35%)	28 (28%)	0.286
Cataracts	146 (80%)	81 (99%)	65 (65%)	<0.001
Better eyes visual acuity, n (%)				
Worse than 20/40	-	21 (26%)	-	
Worse than 20/200	-	3 (4%)	-	
Self-reported better eyes vision, n (%)				
Not sure	-	-	6 (6%)	
Not meeting the requirement for driving	-	-	8 (8%)	
Legal blindness	-	-	5 (5%)	

*We used t-test for age and Wilcoxon rank-sum test for time since diagnosis. For categorical variables, we used chi-squared test or Fisher's exact test, where appropriate.

IQR = Interquartile range

JHU = Wilmer Eye Institute, Johns Hopkins University

MUST FS = Multicenter Uveitis Steroid Treatment Trial Follow-up Study

PENN = Sheie Eye Institute, University of Pennsylvania

SD = Standard deviation

Table 3-3. Estimates of treatment outcome importance using best-worst scaling (BWS)

Treatment outcomes	Total counts in 10 BWS tasks across participants, n (%)*			Aggregate BWS scores	Individual BWS scores		Standardized individual BWS scores	
	Most worrying	Least worrying	Not chosen		Median	Interquartile range	Median	Interquartile range
Vision not meeting the requirement for driving	506 (56%)	176 (19%)	228 (25%)	330	3	0 - 5	80	50 - 100
Glaucoma	481 (53%)	94 (10%)	335 (37%)	387	2	1 - 4	70	60 - 90
Needing eye surgery	385 (42%)	137 (15%)	388 (43%)	248	1	0 - 3	60	50 - 80
Needing medicine for high blood pressure/cholesterol	158 (17%)	471 (52%)	281 (31%)	-313	-2	-4 - 0	30	10 - 50
Cataracts	113 (12%)	446 (49%)	351 (39%)	-333	-2	-3 - -1	30	20 - 40
Infection (e.g., sinusitis)	178 (20%)	496 (55%)	236 (26%)	-318	-3	-5 - 0	20	0 - 50

* N=910, the number of times that an object appeared across 10 BWS tasks and 182 participants

Table 3-4. Associations between patient characteristics and standardized individual best-worst scaling (BWS) scores of each treatment outcome

Bivariate analysis						
	Standardized individual BWS scores of treatment outcomes					
Characteristics	Vision not meeting the requirement for driving	Cataracts	Glaucoma	Needing eye surgery	Needing medicine for high blood pressure/cholesterol	Infection (e.g., sinusitis)
	Mean difference in scores (95% confidence interval)					
MUST FS vs PENN or JHU						
PENN or JHU				*Ref*		
MUST FS	7.9 (-1.8, 17.6)	-3.2 (-9.5, 3.2)	-2.6 (-9.3, 4.2)	-2.5 (-8.8, 3.8)	0.4 (-7.7, 8.4)	0.1 (-9.5, 9.6)
Gender						
Female				*Ref*		
Male	5.1 (-5.9, 16.2)	3.2 (-4.0, 10.3)	-4.7 (-12.4, 2.9)	2.2 (-4.9, 9.3)	-9.6 (-18.6, -0.6)	3.8 (-7.0, 14.6)
Age						
1 year difference	0.2 (-0.1, 0.5)	-0.3 (-0.5, -0.1)	0.1 (-0.1, 0.3)	-0.1 (-0.3, 0.1)	0.1 (-0.2, 0.3)	0.0 (-0.3, 0.3)
Race						
Other				*Ref*		
White	**21.0 (11.6, 30.4)***	-8.0 (-14.2, -1.6)	-0.1 (-6.9, 6.8)	-2.0 (-8.3, 4.3)	0.2 (-7.9, 8.3)	-11.3 (-20.8, -1.9)
Education						
Grade 9 to12				*Ref*		
Some college or higher	**14.5 (4.6, 24.5)***	2.8 (-3.8, 9.4)	3.8 (-3.2, 10.7)	0.1 (-6.4, 6.6)	-4.5 (-12.8, 3.8)	**-16.9 (-26.4, -7.3)***
Time since diagnosis						
1 year difference	0.1 (-0.5, 0.7)	-0.3 (-0.6, 0.1)	-0.2 (-0.6, 0.2)	0.2 (-0.2, 0.5)	0.2 (-0.3, 0.7)	0.0 (-0.6, 0.6)
Location of uveitis						
Other				*Ref*		
Anterior	-10.9 (-21.5, -0.2)	3.3 (-3.7, 10.3)	4.6 (-2.8, 12.0)	2.1 (-4.8, 9.0)	4.4 (-4.4, 13.1)	-3.6 (-14.0, 6.9)
Had cataracts						
No	-	*Ref*	-	-	-	-
Yes	-	-6.5 (-14.3, 1.4)	-	-	-	-

Had glaucoma						
No	-	-	*Ref*	-	-	-
Yes	-	-	-5.1 (-12.4, 2.2)	-	-	-
Had eye surgery						
No	-	-	-	*Ref*	-	-
Yes	-	-	-	1.8 (-5.5, 9.2)	-	-
Had high blood pressure or cholesterol						
No	-	-	-	-	*Ref*	-
Yes	-	-	-	-	2.8 (-5.2, 10.9)	-

* P-values < 0.008

MUST FS = Multicenter Uveitis Steroid Treatment Trial Follow-up Study

PENN = Sheie Eye Institute, University of Pennsylvania

JHU = Wilmer Eye Institute, Johns Hopkins University

Table 3-5. Associations (adjusted) between patient characteristics and standardized individual best-worst scaling (BWS) scores of each treatment outcome

Multivariate analysis[1]						
	Standardized individual BWS scores of treatment outcomes					
Characteristics	Vision not meeting the requirement for driving	Cataracts	Glaucoma	Needing eye surgery	Needing medicine for high blood pressure/cholesterol	Infection (e.g., sinusitis)
	Mean difference in scores (95% confidence interval)					
MUST FS vs PENN or JHU						
PENN or JHU				*Ref*		
MUST FS	6.2 (-5.6, 17.9)	1.3 (-7.1, 9.7)	-1.3 (-10.0, 7.4)	-3.3 (-12.1, 5.4)	0.0 (-10.4, 10.3)	-2.6 (-14.5, 9.3)
Gender						
Female				*Ref*		
Male	8.3 (-2.6, 19.1)	2.1 (-5.3, 9.4)	-5.4 (-13.4, 2.6)	1.4 (-6.1, 9.0)	-10.0 (-19.5, -0.4)	3.3 (-7.7, 14.3)
Age						
1 year difference	0.3 (0.0, 0.6)	-0.2 (-0.5, 0.0)	0.1 (-0.2, 0.3)	-0.2 (-0.4, 0.1)	0.0 (-0.3, 0.3)	0.0 (-0.3, 0.3)
Race						
Other				*Ref*		
White	**16.7 (7.0, 26.4)**[2]	-8.0 (-14.6, -1.4)	1.9 (-5.3, 9.1)	-1.8 (-8.6, 5.0)	2.3 (-6.3, 10.8)	-10.9 (-20.8, -1.1)
Education						
Grade 9 to12				*Ref*		
Some college or higher	**15.1 (5.1, 25.1)**[2]	3.1 (-3.7, 9.9)	1.7 (-5.8, 9.1)	1.1 (-5.9, 8.1)	-6.1 (-14.9, 2.8)	**-14.7 (-24.8, -4.5)**[2]
Time since diagnosis						
1 year difference	0.0 (-0.6, 0.6)	-0.1 (-0.5, 0.3)	-0.2 (-0.6, 0.3)	0.2 (-0.2, 0.6)	0.2 (-0.3, 0.7)	-0.1 (-0.7, 0.5)
Location of uveitis						
Other				*Ref*		
Anterior	-6.6 (-19.2, 6.1)	1.6 (-7.0, 10.2)	4.7 (-4.6, 14.1)	1.2 (-7.6, 10.1)	6.2 (-4.9, 17.4)	-6.9 (-19.7, 6.0)
Had cataracts						
No	-	*Ref*	-	-	-	-

Yes	-	-4.1 (-13.1, 4.9)	-	-	-	-
Had glaucoma						
No	-	-	*Ref*	-	-	-
Yes	-	-	-3.2 (-10.8, 4.4)	-	-	-
Had eye surgery						
No	-	-	-	*Ref*	-	-
Yes	-	-	-	4.1 (-4.9, 13.1)	-	-
Had high blood pressure or cholesterol						
No	-	-	-	-	*Ref*	-
Yes	-	-	-	-	1.7 (-7.2, 10.6)	-

[1] Six multiple linear regression models (each with one of the six standardized individual BWS scores as the dependent variable) were constructed. In each model, we adjusted for all variables listed in each column.

[2] P-values < 0.008

MUST FS = Multicenter Uveitis Steroid Treatment Trial Follow-up Study

PENN = Sheie Eye Institute, University of Pennsylvania

JHU = Wilmer Eye Institute, Johns Hopkins University

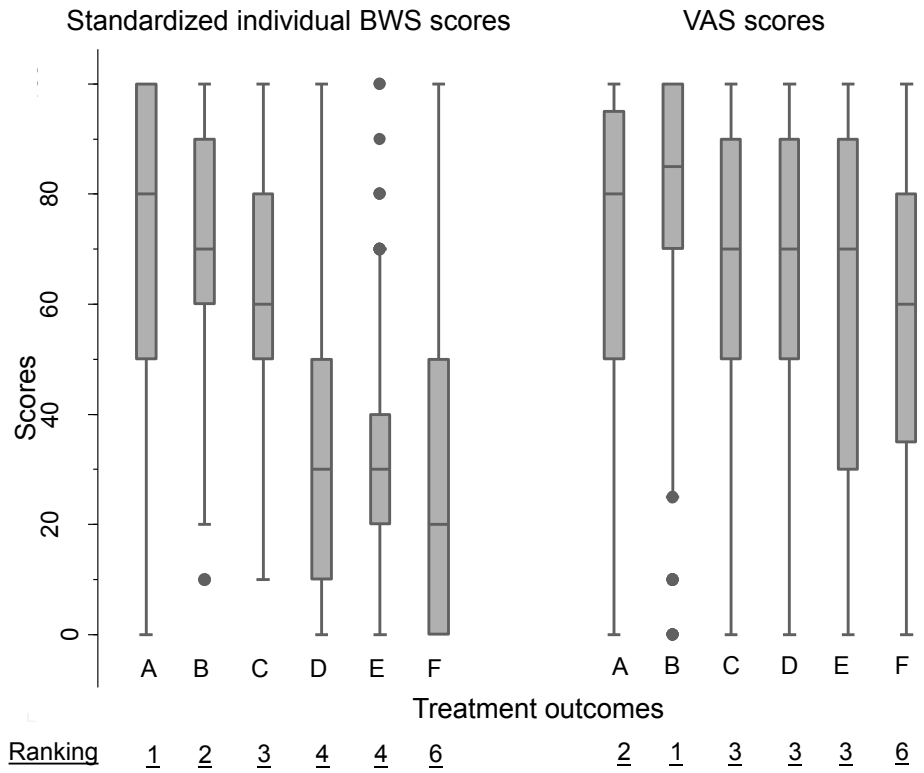


Figure 3-1. Standardized individual best-worst scaling (BWS) scores, visual analogue scale (VAS) scores and rankings for the treatment outcomes.

The box plots represent the median, interquartile range and 95% confidence interval for the scores of each outcome. Rankings are based on medians of the scores. A = vision not meeting the requirement for driving; B = glaucoma; C = needing eye surgery; D = needing medicine for high blood pressure/cholesterol; E = cataracts; F = infection (e.g., sinusitis).

Appendix 3-1. Copy of Patient Preferences Questionnaire

Patient Preferences Questionnaire

The goal of this survey is to help us understand the treatment preferences of patients with uveitis.

Please read and follow the instructions carefully.

Thank you for completing this survey!

Part 1

Below is an example of a question in part 1.

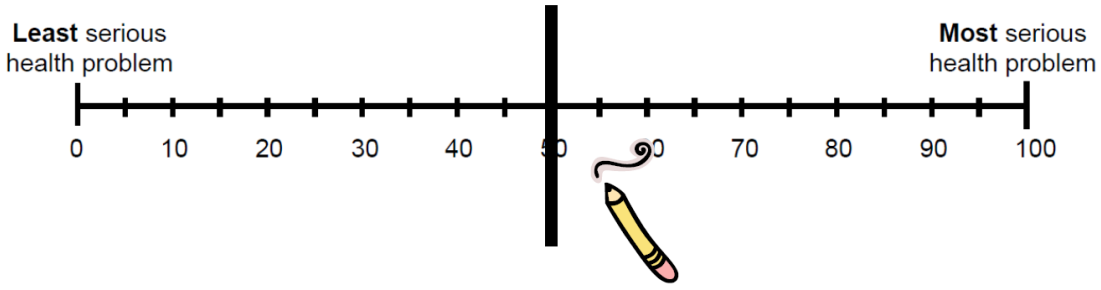
How serious do you think the following health problem is?

Migraine headaches

Migraine headaches are very painful headaches that tend to recur. Sometimes they are associated with symptoms such as vomiting, nausea and seeing an aura.

Please indicate this by drawing a line on the point of the scale.

0 = Least serious health problem; 100 = Most serious health problem

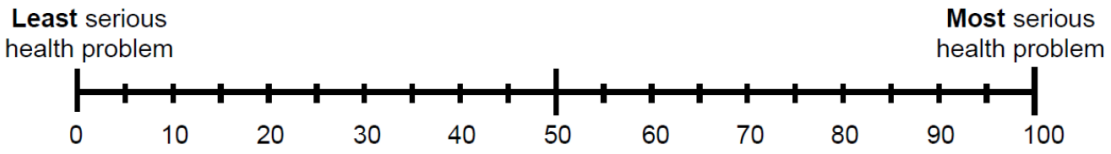


Start here:

Please rate the health problem in each question by drawing a line on the scale.

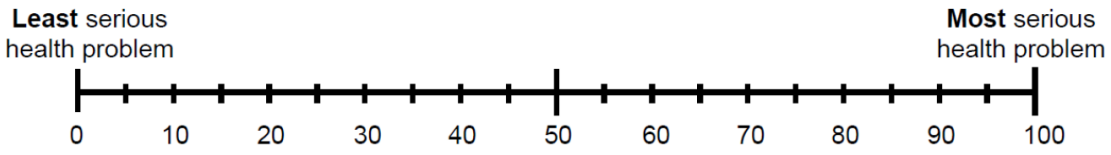
1. How serious do you think the following health problem is?

Vision does not meet the requirement for driving
20/20 is a way to express normal vision. This means that at 20 feet you can see clearly what should be seen at that distance. To obtain a driver's license, your vision with glasses correction must be at least 20/40 (meaning that the letters you can read at 20 feet away can be read correctly by people with normal vision at 40 feet away). You cannot drive on public roads if your vision does not meet this requirement.



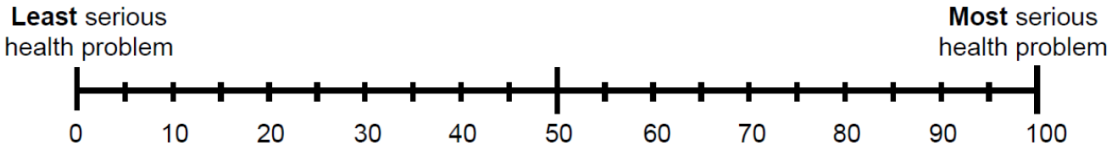
2. How serious do you think the following health problem is?

Cataracts
A cataract is a clouding of the lens in the eye. Patients may not notice that they have cataracts at first. As the eye lens becomes more and more clouded, vision becomes impaired and may result in blurry vision. Other symptoms of cataracts include having poor night vision or seeing halos around lights. Cataract surgery may be recommended by your doctor. It is usually safe and successful in restoring vision.



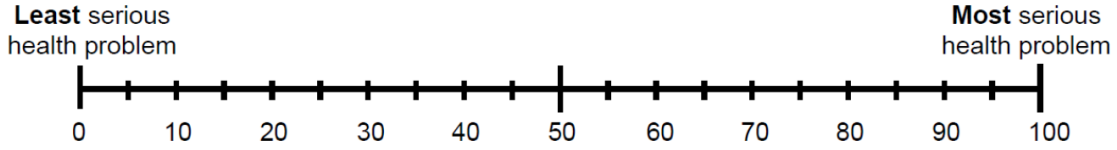
3. How serious do you think the following health problem is?

Glaucoma
Glaucoma is an eye disease that can lead to vision loss. Over time patients with glaucoma can lose side vision and have reduced ability to see shades of gray. Moderate glaucoma can lead to reduced mobility, slower and more difficult reading. Treatments include medications, like daily eye drops, and sometimes surgery. There is no cure for glaucoma. Once someone has it they will probably have to take medications for it for the rest of their life unless they have surgery.



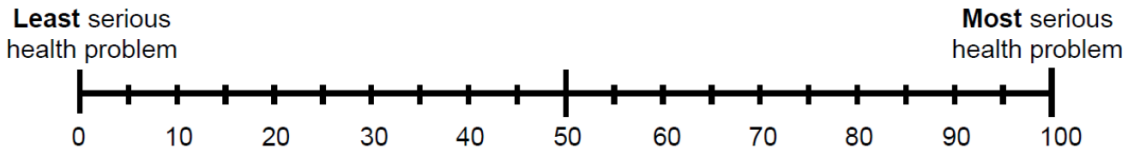
4. How serious do you think the following health problem is?

Needing eye surgery (for cataracts or high eye pressure)
Patients sometimes develop problems with their eyes that need surgery, for example, a cataract. Cataract surgery is done with a painless operation that takes about 30 minutes. It does not require you to stay overnight in the hospital. Other patients may need surgery to lower the pressure in their eyes. This type of surgery takes about 45 minutes. You can go home the same day just as with cataract surgery. Both surgeries usually work well, but any eye surgery has risks including eye infection which can rarely lead to vision loss.



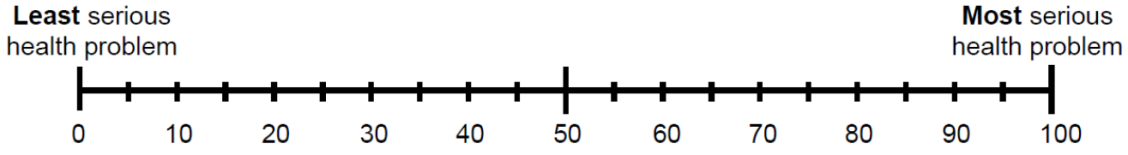
5. How serious do you think the following health problem is?

Needing medicine for controlling high blood pressure or cholesterol
High blood pressure and high cholesterol (fats in the blood) can lead to heart attack or stroke. Keeping your blood pressure and cholesterol under control can reduce your risk of heart disease. Many medicines can help you control blood pressure or cholesterol. For example, if you have high blood pressure, doctors may ask you to take antihypertensive drugs. If you have high cholesterol, they may ask you to take a statin drug. There can be side effects of these drugs such as headache and fatigue but they are generally mild.



6. How serious do you think the following health problem is?

Infection
Some drugs work by suppressing your immune system, so after taking the drugs you may be at a higher risk of infections. For example, sinusitis is one type of infection commonly seen. When you have sinusitis, your nose is stuffy. You may cough, feel tired, get a headache or have a fever. Doctors may ask you to take antibiotics to control the infection. Usually infections resolve quickly after they are treated.



Part 2

Below is an <u>example</u> of a question in part 2.

What are the health problems that would most and least worry you?

Please look at the set of health problems below.

Please indicate:

- The health problem that would worry you **most** (check only one)
- The health problem that would worry you **least** (check only one)

<u>Most</u> worrying	**Health problem**	**<u>Least</u> worrying**
(**X** 1)	a) Feeling tired	(2)
(1)	b) Losing weight	(**X** 2)
(1)	c) Having a runny nose	(2)

The person who answered this question was <u>most</u> worried about "feeling tired" and <u>least</u> worried about "losing weight".

Start here

What are the health problems that would **most and least** worry you?

Please look at the sets of health problems below.

For each set please indicate:

- The health problem that would worry you **most** (check only one)
- The health problem that would worry you **least** (check only one)

7.

What are the health problems that would **most and least** worry you?

Most worrying *check only one*	**Health problem**	**Least** worrying *check only one*
(1)	a) Needing eye surgery	(2)
(1)	b) Needing medicine for high blood pressure/cholesterol	(2)
(1)	c) Decreased vision (not meeting the requirement for driving)	(2)

8.

What are the health problems that would **most and least** worry you?

Most worrying *check only one*	**Health problem**	**Least** worrying *check only one*
(1)	a) Diagnosed with cataracts	(2)
(1)	b) Diagnosed with glaucoma	(2)
(1)	c) Decreased vision (not meeting the requirement for driving)	(2)

9.

What are the health problems that would **most and least** worry you?

Most worrying *check only one*	**Health problem**	**Least** worrying *check only one*
(1)	a) Infection (e.g., sinusitis)	(2)
(1)	b) Decreased vision (not meeting the requirement for driving)	(2)
(1)	c) Diagnosed with cataracts	(2)

10.

What are the health problems that would **most and least** worry you?

Most worrying *check only one*	**Health problem**	**Least** worrying *check only one*
(1)	a) Decreased vision (not meeting the requirement for driving)	(2)
(1)	b) Needing medicine for high blood pressure/cholesterol	(2)
(1)	c) Diagnosed with glaucoma	(2)

11.

What are the health problems that would **most and least** worry you?

Most worrying *check only one*	**Health problem**	**Least** worrying *check only one*
(1)	a) Decreased vision (not meeting the requirement for driving)	(2)
(1)	b) Needing eye surgery	(2)
(1)	c) Infection (e.g., sinusitis)	(2)

12.

What are the health problems that would **most and least** worry you?

Most worrying *check only one*	**Health problem**	**Least** worrying *check only one*
(1)	a) Diagnosed with glaucoma	(2)
(1)	b) Diagnosed with cataracts	(2)
(1)	c) Needing eye surgery	(2)

13.

What are the health problems that would **most and least** worry you?

Most worrying *check only one*	**Health problem**	**Least** worrying *check only one*
(1)	a) Diagnosed with glaucoma	(2)
(1)	b) Infection (e.g., sinusitis)	(2)
(1)	c) Needing medicine for high blood pressure/cholesterol	(2)

14.

What are the health problems that would **most and least** worry you?

Most worrying *check only one*	**Health problem**	**Least** worrying *check only one*
(1)	a) Needing medicine for high blood pressure/cholesterol	(2)
(1)	b) Needing eye surgery	(2)
(1)	c) Diagnosed with cataracts	(2)

15.

What are the health problems that would **most and least** worry you?

Most worrying *check only one*	**Health problem**	**Least** worrying *check only one*
(1)	a) Infection (e.g., sinusitis)	(2)
(1)	b) Diagnosed with cataracts	(2)
(1)	c) Needing medicine for high blood pressure/cholesterol	(2)

16.

What are the health problems that would **most and least** worry you?

Most worrying *check only one*	**Health problem**	**Least** worrying *check only one*
(1)	a) Needing eye surgery	(2)
(1)	b) Infection (e.g., sinusitis)	(2)
(1)	c) Diagnosed with glaucoma	(2)

Part 3 - Questions about you

17. What is your gender? Please check one.

Male ..... (1)

Female ..... (2)

18. Are you of Hispanic or Latino origin? Please check one.

Yes ..... (1)

No ..... (2)

19. What is your race? Please check one.

American Indian/Alaskan Native ..... (1)

Asian ..... (2)

Native Hawaiian or other Pacific

Islander ..... (3)

Black or African American ..... (4)

White ..... (5)

Other ..... (6)

Please specify

_____

20. What is the highest level of education that you have attained? Please check one.

No formal education ..... (1)

Grade 8 or less ..... (2)

Grade 9, 10, or 11 ..... (3)

Grade 12 or high school graduate (4)

Some college (5)

College graduate (6)

One or more years post college (7)

21. What is your current employment status? Please check one.

Disabled (unable to work) (1)

Employed with income (2)

Homemaker (3)

Retired (4)

Student (5)

Unemployed (6)

22. What is your insurance status?

Please check all that apply.

a) Uninsured (1)

b) Medicare (1)

c) Medicaid (1)

d) Veterans Administration (1)

e) Private health Insurance (1)

23. What is your age?

___ ___ years old

24. Around what age were you diagnosed with uveitis?

__ __ years old

25. Which of your eyes is affected by uveitis? Please check one.

Right eye.....(1)

Left eye.....(2)

Both eyes.....(3)

Not sure.....(4)

26. What is the visual acuity of your better-seeing eye? Please check one.

20/40 or better .....(1)

Worse than 20/40 but better than

20/200.....(2)

20/200 or worse.....(3)

Not sure.....(4)

27. What treatments have been used specifically to manage your uveitis?

Please check all that apply.

a) I have not received any treatments.....(1)

b) Corticosteroid eye drops.....(1)

c) Corticosteroid injections.....(1)

d) Surgical corticosteroid implant .....(1)

e) Oral corticosteroids.....(1)

f) Immunosuppressive drug therapy.....(1)

g) Biologics.....(1)

h) Other.....(1)

Please specify

_____

i) Not sure.....(1)

28. Have you been diagnosed with any of the following diseases?

Please check all that apply.

a) Behcet's disease.....(1)

b) Sarcoidosis.....(1)

c) Multiple sclerosis.....(1)

d) Juvenile rheumatoid arthritis.....(1)

e) Familial systemic juvenile granulomatosis.....(1)

f) None of the above.....(1)

g) Not sure.....(1)

29. Have you had any of the following health problems before?

Please check all that apply.

a) Glaucoma.....(1)

b) Cataracts.....(1)

c) Visual acuity 20/200 or worse (legal blindness).....(1)

d) Eye surgery.....(1)

e) Any eye complications after eye surgery .....(1)

f) None of the above .....(1)

g) Not sure .....(1)

30. Have you had any of the following health problems before?

Please check all that apply.

a) High blood pressure .....(1)

b) Heart disease .....(1)

c) Elevated cholesterol .....(1)

d) Diabetes .....(1)

e) Osteoporosis .....(1)

f) Fracture .....(1)

g) Systemic infection .....(1)

h) None of the above .....(1)

i) Not sure .....(1)

Additional analysis

Table 3-6. Associations between patient characteristics and standardized individual best-worst scaling (BWS) scores of each treatment outcome (non-parametric test)[1]

	Standardized individual BWS scores of treatment outcomes					
	Vision not meeting the requirement for driving	Cataracts	Glaucoma	Needing eye surgery	Needing medicine for high blood pressure/cholesterol	Infection (e.g., sinusitis)
Characteristics	P-values					
MUST FS vs PENN or JHU	0.022	0.410	0.386	0.363	0.771	0.960
Gender (Female vs Male)	0.409	0.525	0.208	0.520	0.056	0.466
Age (Spearman correlation)	0.311 (Spearman's rho = 0.076)	0.012 (Spearman's rho = -0.187)	0.226 (Spearman's rho = 0.091)	0.151 (Spearman's rho = -0.108)	0.967 (Spearman's rho = 0.003)	0.431 (Spearman's rho = 0.059)
Race (White vs Other)	**0.000**[2]	**0.003**[2]	0.970	0.439	0.775	**0.007**[2]
Education (Some college or higher vs Grade 9 to 12)	0.016	0.612	0.228	0.972	0.324	**0.000**[2]
Time since diagnosis (Spearman correlation)	0.346 (Spearman's rho = 0.071)	0.095 (Spearman's rho = -0.126)	0.300 (Spearman's rho = -0.078)	0.946 (Spearman's rho = 0.005)	0.254 (Spearman's rho = 0.086)	0.938 (Spearman's rho = 0.006)
Location of uveitis (Anterior vs Other)	0.015	0.657	0.248	0.430	0.554	0.659

Had cataracts (Yes vs No)	-	0.084	-	-	-	-
Had glaucoma (Yes vs No)	-	-	0.144	-	-	-
Had eye surgery (Yes vs No)	-	-	-	0.911	-	-
Had high blood pressure or cholesterol (Yes vs No)	-	-	-	-	0.838	-

[1]We used Wilcoxon rank-sum test when the characteristics variable is categorical and calculated Spearman rank correlation when the characteristics variable is continuous.

[2]P-values < 0.008

MUST FS = Multicenter Uveitis Steroid Treatment Trial Follow-up Study

PENN = Sheie Eye Institute, University of Pennsylvania

JHU = Wilmer Eye Institute, Johns Hopkins University

CHAPTER FOUR

USING A PATIENT-CENTERED APPROACH TO BENEFIT-HARM ASSESSMENT IN TREATMENT DECISION-MAKING: A CASE STUDY OF CORTICOSTEORID THERAPY IN UVEITIS

ABSTRACT

Purpose

Interpreting findings from comparative effectiveness trials sometimes can be difficult since multiple treatment outcomes with different importance to patients are at play. The goal of this study was to demonstrate an approach to conducting quantitative benefit-harm assessment of competing treatments in which patient preferences are incorporated.

Methods

We conducted a benefit-harm assessment using data from the Multicenter Uveitis Steroid Treatment Trial that compared corticosteroid implant versus systemic corticosteroids in treating non-infectious intermediate, posterior, and panuveitis. We focused on patient-centered treatment outcomes and incorporated measures of patient preferences that were derived from a recent survey into our estimation of the benefit-harm balance over time. We also examined the joint occurrence of benefit and harm outcomes in the trial participants.

Results

Benefit-harm metrics were calculated for each time point that summarized the numbers of cases with each outcome, caused or prevented by the implant therapy compared with systemic therapy if 1000 patients were treated. The benefit-harm metric was -129 (95% CI: -242 to -14), -317 (-436 to -196), -390 (-514 to -264) and -526 (-687

to -368) and the probability of the metric being positive was 1%, 0%, 0%, and 0% at 6, 12, 18, and 24 months follow-up, respectively, which implies that implant therapy has a worse benefit-harm balance than systemic therapy. Our interpretation of the benefit-harm balance remained similar in the analyses when we varied the weights assigned to outcomes and when we considered the joint occurrence of benefits and harms.

Conclusions

We demonstrated a quantitative approach to benefit-harm assessment that is focused on patient-centered outcomes and that incorporates patient preferences. The approach estimates the benefit-harm balance between competing treatments and can be helpful for making treatment decisions.

INTRODUCTION

Randomized comparative effectiveness trials are used to make head-to-head comparisons between competing treatments and generate evidence on treatment effectiveness and safety.[1] For example, the Multicenter Uveitis Steroid Treatment (MUST) Trial was designed to compare two active treatments that both work in treating non-infectious intermediate, posterior, and panuveitis.[2] Eligible patients were randomized to either fluocinolone acetonide implant (implant therapy) or systemic corticosteroids plus immunosuppression when indicated (systemic therapy) and followed for 24 months. Although no significant difference in visual acuity change from baseline was detected between the two treatment groups, implant therapy was associated with more ocular adverse effects (e.g., cataracts and glaucoma) while systemic therapy was associated with more systemic adverse effects (e.g., hypertension and infections).[3]

Interpreting findings from such a comparative effectiveness trial is difficult for most patients, clinicians, or policy makers. It often remains uncertain if one treatment is superior to the other because multiple benefit and harm outcomes are at play. A number of quantitative approaches for benefit-harm assessment exist that deal with such multidimensional tasks and have the potential to substantially aid our interpretation and judgment of comparative treatment effectiveness.[4, 5] Quantitative benefit-harm assessments can consider multiple outcomes and patient characteristics to provide a single metric, if desired, that reflects the benefit-harm balance. Since patients may

perceive the importance of each benefit or harm outcome differently, these summary metrics of benefits and harms also allow for taking into account the patient preferences for treatment outcomes.[5]

Joint occurrence of benefit and harm outcomes is another important, though usually neglected, issue to consider when doing quantitative benefit-harm assessment.[5-9] Occurrence of outcomes can be correlated rather than independent if, for instance, they typically co-occur with each other within patients. Nonetheless, treatment effects on each outcome are traditionally reported separately in the medical literature.[5, 10] Some approaches to benefit-harm assessment that account for such joint occurrence of outcomes have been developed, yet are not commonly applied.[11-15]

The objective of the present study was to demonstrate a patient-centered approach to benefit-harm assessment that specifically incorporates patient preferences. We used the example of the MUST Trial that compared implant therapy with systemic therapy in patients with intermediate, posterior and panuveitis. With patient-level data from the trial, we also examined the joint occurrence of benefit and harm outcomes in these patients.

METHODS

Data Source: The MUST Trial

The MUST Trial is a randomized comparative effectiveness trial designed to compare corticosteroid implant versus systemic corticosteroids in patients with non-infectious intermediate, posterior and panuveitis. Details on the study design and primary results of the trial were reported previously.[2, 3] In brief, patients aged 13 or older who had uveitis in at least one eye and who were indicated for systemic corticosteroids were randomized to implant or systemic therapy. Patients in the implant therapy group received in the eligible eye a surgical fluocinolone acetonide implant (0.59 mg) that delivers corticosteroids intravitreally. Patients in the systemic therapy group were treated with oral corticosteroids (prednisone) supplemented with immunosuppressive agents according to treatment guidelines.[2, 16] Participants were recruited from 23 clinical centers across the United States, the United Kingdom, and Australia from December 2005 to December 2008 and were followed for at least 24 months. All participants provided their informed consent and the institutional review boards at the clinical centers and resource centers approved the study.

Outcomes Included for Benefit-Harm Assessment

Our aim was to do a patient-centered benefit-harm assessment, so we focused on patient-centered outcomes (as opposed to biomarkers or other surrogate outcomes)

measured in the MUST Trial that are deemed important to decision-making. The goal of treating these patients is to preserve their vision from getting worse and, at the same time, to minimize adverse effects caused by the treatments. For the vision outcome, we examined the proportion of patients who did not respond to treatment at each time point, which was defined as their best-corrected visual acuity in the better-seeing eye staying at or even decreasing to worse than 20/40 (vision not meeting the requirement for driving). For ocular and systemic adverse effects of treatments, we examined at each time point the proportion of patients who had the following outcomes: incident cataracts, incident glaucoma, requiring cataract surgery, requiring intraocular pressure-lowering surgery (glaucoma surgery), prescription-requiring hypertension, prescription-requiring hyperlipidemia, and prescription-requiring infections. The time frames of our assessment of these outcomes were 6, 12, 18, and 24 months after randomization. The unit of the analysis was the individual patient.

Relative Importance for Outcomes

In a benefit-harm assessment to compare multiple outcomes on a single metric, we needed to assign weights to outcomes based on their relative importance. Therefore, in our analysis we assigned the weights using data from a recent survey that elicited patient preferences for treatment outcomes in non-infectious uveitis.[17] Briefly, the survey asked 182 patients with non-infectious uveitis to complete the "best-worst scaling" exercise, a preference-elicitation method used in health economics where

patients have the tasks to repeatedly select the most and the least worrying from a list of outcomes (in the survey 3 outcomes per task). Among all outcomes being compared, vision not meeting the requirement for driving, incident glaucoma, requiring cataract surgery and requiring glaucoma surgery were thought to be more worrying by respondents as compared against incident cataracts, prescription-requiring hypertension, prescription-requiring hyperlipidemia, or prescription-requiring infections. Estimates of patient preferences for these outcomes were incorporated into our benefit-harm assessment.

Benefit-Harm Metric

We summarized the treatment effects on different outcomes (weighted by each outcome's relative importance) in a "benefit-harm metric" that reflects the benefit-harm balance. First, based on the MUST Trial data, we calculated the numbers of cases with respective outcome (outcome x) if 1000 patients were treated with implant therapy ($N_{X,IMP}$) or systemic therapy ($N_{X,SYS}$). Second, we calculated the cases prevented or caused if 1000 patients were treated with implant therapy as compared against systemic therapy ($N_x = N_{X,SYS} - N_{X,IMP}$). A positive number represents the number of cases prevented and a negative number represents the number of cases caused by implant therapy. Third, we assigned weights (W_X, relative importance) to these outcomes according to the preference measures elicited in the patient preferences survey. We then computed a benefit-harm metric that summarizes the overall numbers caused or

prevented by implant therapy and that incorporates the relative importance of outcomes. If the benefit-harm metric is positive, it suggests that implant therapy is superior to systemic therapy since the implant therapy prevented more cases overall.

We computed the benefit-harm metrics at 6, 12, 18, and 24 months after randomization. In addition, we varied the weights assigned to outcomes as sensitivity analyses to evaluate whether our study conclusions would change with regards to different assigned weights. We used bootstrapping approach to incorporate the statistical uncertainty. Bootstrapping was conducted by drawing a random sample from the observations in each treatment group and calculating the benefit-harm metric 10000 times. Thereby, we obtained 10000 replicates of the metric to compute its 95% confidence interval (CI) and the probability that the metric is positive. Analyses were performed using Stata 11.2 (StataCorp LP, Colllege Station, TX) and R statistical software 3.0.1.

Joint Occurrence of Benefits and Harms

To examine the joint occurrence of benefit and harm outcomes in patients in the MUST Trial, we defined two benefit categories (based on patients' vision outcome) and three harm categories (based on their experiences with adverse effects). The "benefit categories" were: (1) patients' visual acuity of the better-seeing eye stayed at or improved

to 20/40 or better ("With benefits"); (2) patients' visual acuity of the better-seeing eye stayed at or decreased to worse than 20/40 ("No benefits"). The "harm categories" were defined as: (1) patients had no adverse events of interest ("No harms"); (2) patients only had the adverse events that are less worrying (according to the preference survey abovementioned) including incident cataracts, prescription-requiring hypertension, prescription-requiring hyperlipidemia, or prescription-requiring infections ("Minor harms"); (3) patients had any of the adverse events that are more worrying including incident glaucoma, requiring cataract surgery and requiring glaucoma surgery ("Moderate harms"). Finally, we created "benefit-harm categories" that consider benefits and harms jointly: two benefit categories x three harm categories, e.g., "With benefits/No harms", "With benefits/Minor harms", "With benefits/Moderate harms", and so on. One more category ("Missing data") was created that included patients with missing data of their visual acuity. These patients were died or loss to follow-up, or their visual acuity was not or could not be measured (see **Table 4-1** for different benefit-harm categories).

We then assigned every patient in the MUST Trial to each of the seven benefit-harm categories based on their experiences with benefit and harm outcomes during follow-up. We compared the distributions between the two treatment groups at each time point. We also did stratified analysis by baseline vision (visual acuity of the better-seeing eye 20/40 or better, or worse than 20/40 at baseline) and we examined the distribution of benefit-harm categories at 24 months.

RESULTS

Table 4-2 shows the data input for the benefit-harm assessment, the proportion of patients in the MUST trial who ever had each outcome during follow-up (by treatment group) and the weights (i.e., reflecting patient preferences) assigned in the main and sensitivity analyses. Patients' outcomes at four different time points are presented: 6, 12, 18, and 24 months after randomization. At baseline, 82% (210/255) of patients had cataracts and 3% (6/237; 18 patients with missing data) had glaucoma, so these patients were considered not at risk when quantifying incident cataracts or incident glaucoma. In the main analysis, the weights assigned were based on the preference measures obtained in the preference-elicitation survey. We varied the weights in the first sensitivity analysis by assigning 1.0 to more worrying outcomes and 0.5 to less worrying outcomes. In sensitivity analysis two, we assigned 1.0 to the visual acuity outcome (as this is the primary outcome in the trial) and 0.5 to other more worrying outcomes and 0.25 to less worrying outcomes.

An example of calculation of the benefit-harm metric (main analysis, 24 months follow-up) is provided in **Table 4-3**. We calculated the numbers of cases with respective outcome if 1000 patients were treated with implant or systemic therapy and then the numbers of cases with respective outcome caused or prevented by implant therapy. We assumed at baseline 82% of patients had already had cataracts and 3% of patients had had glaucoma (based on data from the MUST Trial). Take the incident glaucoma outcome as

an example. Our calculations show that there would be 226 and 60 cases with incident glaucoma if 1000 patients were treated with implant and systemic therapy, respectively. Thus, 166 cases with incident glaucoma would be in excess comparing implant therapy with systemic therapy. The numbers caused or prevented by implant therapy were then multiplied by the weights, and were summed to compute the benefit-harm metric.

Results of the main and sensitivity analyses of the benefit-harm metric at each time point are shown in **Table 4-4**. In the main analysis, the benefit-harm metric is -129, -317, -390 and -526 at 6, 12, 18, and 24 months follow-up, respectively, implying that implant therapy has a worse benefit-harm balance than systemic therapy. The 95% CIs and the probability that the metric is positive, meaning that implant therapy would be superior to systemic therapy, is 1%, 0%, 0%, and 0%, respectively. Results of the sensitivity analyses are similar. The benefit-harm metrics are more and more distant from 0 (negative) across the 6, 12, 18, and 24 months follow-up, and the probabilities of the index being positive are all small or 0%. This suggests that systemic therapy may be superior to implant therapy given the outcomes and time frames defined in our assessment.

Table 4-5 shows the patients' outcomes at 24 months follow-up by accounting for the joint occurrence of benefits and harms. At 24 months follow-up, 66% of patients in implant group versus 71% in systemic group were in the "With benefits" category.

Combined with their experience with harms, most patients (49%) in the implant group were assigned to the "With benefits/Moderate harms" category, while only 23% in systemic group were in this category.

Results stratified by baseline vision are also shown in **Table 4-5**. Most patients started with vision 20/40 or better of their better eye managed to maintain their vision (82% in both treatment groups) at 24 months follow-up. On the other hand, among patients started with vision worse than 20/40 of their better eye, 36% of patients in implant group versus 43% in systemic group got improved in their vision to 20/40 or better. Generally speaking, regardless of the benefit category they were assigned to, patients in the implant group were more likely to have moderate harms and patients in systemic group were more likely to have no harms or minor harms. We also plotted the distribution of patients to assigned benefit-harm categories by treatment group over time in the **Figure 4-1**.

DISCUSSION

Based on a randomized comparative effectiveness trial, we quantitatively assessed the benefits and harms of corticosteroid implant versus systemic corticosteroids in patients with non-infectious intermediate, posterior, and panuveitis. We adopted a patient-centered approach in our assessment where we focused on patient-centered

outcomes and assigned weights (patient preferences for these outcomes indicating their relative importance) to each outcome according to a preference-elicitation survey of patients. Given the outcomes and time frames we defined, the results showed that systemic therapy may be superior to implant therapy.

Using a *quantitative* approach to benefit-harm assessment, as we did in this study may provide more clarity and transparency than using a *qualitative* approach, in particular of the decision-making being challenging due to many different outcomes. In a quantitative benefit-harm assessment, a common metric is often needed to summarize and compare the treatment effects for both benefits and harms.[5] Treatment effects in clinical trial reports are commonly expressed as relative risks (e.g., risk ratio, odds ratio, or hazard ratio). But because the same relative risk can translate into considerably different effects for different patients as the respective absolute risks are different, it seems necessary to use absolute risks to put multiple outcomes on the same metric.[4] For example, a relative risk of 0.5 (50% risk reduction) for heart attack can mean that 25% (absolute risk of the control group: 50%) or 1% (absolute risk of the control group: 2%) of the events are prevented. It is thus probably not sensible if we rely on relative risks alone to do a quantitative benefit-harm assessment.

In our analysis, we created a common metric (number of cases with the outcome if 1000 patients were treated) and put all outcomes on the same metric based on the

absolute risks (probabilities). We calculated the numbers of cases with the outcome that would be caused or prevented by the implant therapy compared with systemic therapy, and then properly assigned weights to these numbers to generate the benefit-harm metric. Based on the metric, we were able to tell which treatment would be superior to the other considering both benefit and harm outcomes. We used bootstrapping methods to calculate the 95% CI of the metric and the probability of the metric being positive, which informed us of the statistical uncertainty around the benefit-harm balance. Our quantitative approach to benefit-harm assessment was transparent in that the specific outcomes considered, weights assigned, and the uncertainty of the data were clearly laid out. This is in contrast to using a qualitative approach in which the outcomes and weights used and the assumptions and judgment made at every step of a benefit-harm assessment are often less clear or transparent to readers.

Our study also demonstrated how the joint occurrence of benefits and harms can be examined when doing benefit-harm assessment. In clinical trials, the evaluation of treatment effect is usually done separately for each outcome.[10] However, it would be of great interest to patients and clinicians if data of the joint impacts of benefit and harm outcomes on the *same* individual are available.[10] For example, we computed and plotted the distribution of benefit-harm categories (in which the joint probability of the occurrence of benefits and harms were examined, see the **Figure 4-1**) by treatment group. We found that the distribution of outcomes was somewhat different in the implant and

systemic groups. At 24 months follow-up the distribution of the harm categories was not the same between the two treatment groups, although the proportions of patients who were assigned to “With benefits” category were similar (66% vs 71%). Most patients in the implant group had moderate harms, but patients in the systemic group were more evenly distributed among no, minor, and moderate harms. Such information on the joint occurrence of benefits and harms is not commonly reported in the current literature of clinical trials, but would be extremely helpful when patients and clinicians desire to make personalized treatment decisions.

Assigning weights (relative importance) to outcomes is no doubt the most controversial part of a benefit-harm assessment and is inevitably subjective. Nevertheless, it remains essential, as in clinical practice, because it is not sensible if all outcomes were considered to be of equal importance. Even if one chooses to use a qualitative approach instead, weighting is still implicitly conducted as some outcomes would be viewed more important than others; thus, the results are sensitive to these “implicit weights”.[6] The weighting of outcomes in our benefit-harm assessment were done from a population perspective, where we used preference data elicited from a patient population. We decided up front to conduct a preference-elicitation survey of patients with non-infectious uveitis to help us choose appropriate weights for each outcome. We were comfortable with using the findings from the survey to determine the weights since we found in the survey that the preference estimates were consistent across different patient groups.[17]

However, the variability of the preference estimates were also found to be rather large between individuals.[17] The advantage of using a quantitative approach to benefit-harm assessment is its transparency and reproducibility. The analyis can easily be repeated and modified if anyone disagrees with the outcomes included or the weights assigned to outcomes. Hence, instead of using weights derived from a population, different weights can be assigned depending on each individual's preferences. In addition, a reproducible approach to benefit-harm assessment greatly facilitates sensitivity analyses, which are essential to assess how much the benefit-harm balance changes if different data inputs are chosen.

Selection of outcomes for benefit-harm assessments is a critical step. One may see it as a limitation that we did not consider "control of inflammation" in our benefit-harm assessment. The fluocinolone acetonide implant to treat non-infectious uveitis was approved by the FDA based on its effect on control of inflammation (rate of recurrence of uveitis) [18, 19], and the goal of the treatment is to control the inflammation in the eye with the hope of preserving patient's vision. Since our benefit-harm assessment directly considered visual acuity as the outcome of primary interest rather than using a marker of inflammation as a surrogate, we saw little reason to include control of inflammation in the analysis. This may, however, neglect some potential benefits provided by the implant as implant therapy was shown to be more effective in reducing uveitis activity.[3] Whether this treatment effect on controlling intraocular inflammation can be translated

into the treatment effect on patient's vision would require further investigation in long-term randomized studies.

It is also interesting to note that our results are somewhat contradictory to the findings of health-related quality of life and health utility. For example at 24 months follow-up, the change from baseline of the EQ-5D health utility index was larger in implant group versus systemic group (treatment effect = 0.04, p-value = 0.06)[3]. The EQ-5D instrument assesses several generic health dimensions including mobility, self-care, usual activities, pain/discomfort and anxiety/depression.[20] Such an instrument is designed to be generalizable across various conditions and treatments and often used to compute quality-adjusted life years in cost-effectiveness studies as a way to capture the overall treatment effect on patients' health.[21] Although we have demonstrated that patients do make trade-offs between different treatment outcomes, it is likely that the impact of some outcomes on patients may not be large enough to affect their overall health status, or that the instruments measuring their health status are not sensitive enough in the specific disease of interest.[22] However, our approach is more specific to the decision-making context and we clearly defined the outcomes for benefit-harm assessment. Thus, we may capture the aspects of benefits and harms that are perhaps missed out by more generic health status assessment tools.

Some limitations of this study were identified. In our decision making context, we only focused on patient-centered outcomes that are common in patients with non-infectious uveitis. Some harm outcomes of implant or systemic therapy (e.g., endophthalmitis or psychiatric disorder), though much less common, may also be of importance when making treatment decisions for individuals. The study time frame is 24 months after randomization, the same time points at which the primary outcome was measured. Therefore, our model was not able to capture the potential systemic complications that may occur years later after treatment. To study harms that are rare and long-term, randomized trials are not the most efficient design and it requires data from observational studies or surveillance. Future methodological research is needed to address the question of how data from observational studies or surveillance can be incorporated into a benefit-harm assessment. Finally, when generating the benefit-harm categories and computing the benefit-harm metrics, we categorized visual acuity outcome and combined different harm outcomes together. This can lead to much information loss but it is inevitable in any benefit-harm assessment to reduce some of the multi-dimensionality in order to arrive at estimates of the benefit-harm balance that facilitates treatment decision-making.

In summary, we conducted a quantitative benefit-harm assessment of implant versus systemic therapy in patients with non-infectious intermediate, posterior, and panuveitis to help us make treatment decisions. We have demonstrated how we selected

the outcomes, considered the joint occurrence of benefits and harms and incorporated the patient preferences for outcomes in the analysis. In line with the recent interest in patient-centered outcomes research,[23, 24] we believe our approach is useful and deserves future replications in doing patient-centered benefit-harm assessments.

REFERENCES

1. IOM (Institute of Medicine). Initial National Priorities for Comparative Effectiveness Research. Washington DC: The National Academies Press; 2009.

2. Multicenter Uveitis Steroid Treatment Trial Research Group, Kempen JH, Altaweel MM, Holbrook JT, Jabs DA, Sugar EA. The multicenter uveitis steroid treatment trial: rationale, design, and baseline characteristics. Am J Ophthalmol. 2010;149(4):550-561.e10.

3. Multicenter Uveitis Steroid Treatment (MUST) Trial Research Group, Kempen JH, Altaweel MM, Holbrook JT, Jabs DA, Louis TA, Sugar EA, Thorne JE. Randomized comparison of systemic anti-inflammatory therapy versus fluocinolone acetonide implant for intermediate, posterior, and panuveitis: the multicenter uveitis steroid treatment trial. Ophthalmology. 2011;118(10):1916-26.

4. Gail MH. Using absolute risks to assess the risks and benefits of treatment. Thorax. 2014. [Epub ahead of print].

5. Puhan MA, Singh S, Weiss CO, Varadhan R, Boyd CM. A framework for organizing and selecting quantitative approaches for benefit-harm assessment. BMC Med Res Methodol. 2012;12:173,2288-12-173.

6. Chuang-Stein C. Quantitative risk/benefit assessment: where are we? Proceedings of the Fourth Seattle Symposium in Biostatistics: Clinical Trials. Springer New York. 2013;119-135.

7. Guo JJ, Pandey S, Doyle J, Bian B, Lis Y, Raisch DW. A review of quantitative risk-benefit methodologies for assessing drug safety and efficacy-report of the ISPOR risk-benefit management working group. Value Health. 2010;13(5):657-66.

8. Ouellet D. Benefit-risk assessment: the use of clinical utility index. Expert Opin Drug Saf. 2010;9(2):289-300.

9. The PROTECT Consortium. Review of methodologies for benefit and risk assessment of medication. http://www.imi-protect.eu/documents/ShahruletalReviewofmethodologiesforbenefitandriskassessmentofmedicationMay2013.pdf. 2012 (last accessed on April 12, 2014).

10. Kraemer HC, Frank E. Evaluation of comparative treatment trials: assessing clinical benefits and risks for patients, rather than statistical effects on measures. JAMA. 2010;304(6):683-4.

11. Chuang-Stein C, Mohberg NR, Sinkula MS. Three measures for simultaneously evaluating benefits and risks using categorical data from clinical trials. Stat Med. 1991;10(9):1349-59.

12. Entsuah R, Gorman JM. Global benefit-risk assessment of antidepressants: venlafaxine XR and fluoxetine. J Psychiatr Res. 2002;36(3):111-8.

13. Frank E, Kupfer DJ, Rucci P, Lotz-Wallace M, Levenson J, Fournier J, Kraemer HC. Simultaneous evaluation of the harms and benefits of treatments in

randomized clinical trials: demonstration of a new approach. Psychol Med. 2012;42(4):865-73.

14. Shaffer ML, Watterberg KL. Joint distribution approaches to simultaneously quantifying benefit and risk. BMC Med Res Methodol. 2006;6:48.

15. Wisniewski SR, Chen CC, Kim E, Kan HJ, Guo Z, Carlson BX, Tran QV, Pikalov A. Global benefit-risk analysis of adjunctive aripiprazole in the treatment of patients with major depressive disorder. Pharmacoepidemiol Drug Saf. 2009;18(10):965-72.

16. Jabs DA, Rosenbaum JT, Foster CS, Holland GN, Jaffe GJ, Louie JS, Nussenblatt RB, Stiehm ER, Tessler H, Van Gelder RN, Whitcup SM, Yocum D. Guidelines for the use of immunosuppressive drugs in patients with ocular inflammatory disorders: recommendations of an expert panel. Am J Ophthalmol. 2000;130(4):492-513.

17. Yu T, et al. Patients with non-infectious uveitis consider impaired vision, glaucoma and need for eye surgery worrying adverse treatment outcomes: results of a best-worst scaling study. In preparation.

18. US Department of Health and Human Services. Food and Drug Administration. Center for Drug Evaluation and Research. Medical reviews. Application number:21-737. 2005.

http://www.accessdata.fda.gov/drugsatfda_docs/nda/2005/021737s000_MedR.pdf (accessed April 28, 2014).

19. Pavesio C, Zierhut M, Bairi K, Comstock TL, Usner DW; Fluocinolone Acetonide Study Group.Ophthalmology. Evaluation of an intravitreal fluocinolone acetonide implant versus standard systemic therapy in noninfectious posterior uveitis. 2010 Mar;117(3):567-75

20. Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997 Nov;35(11):1095-108.

21. Siegel JE, Weinstein MC, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses. Panel on Cost-Effectiveness in Health and Medicine. JAMA. 1996 Oct 23-30;276(16):1339-41.

22. Tervonen T, van Valkenhoef G, Buskens E, Hillege HL, Postmus D. A stochastic multicriteria model for evidence-based decision making in drug benefit-risk analysis. Stat Med. 2011 May 30;30(12):1419-28.

23. Methodology Committee of the Patient-Centered Outcomes Research Institute (PCORI). Methodological standards and patient-centeredness in comparative effectiveness research: the PCORI perspective. JAMA. 2012 Apr 18;307(15):1636-40.

24. Gagnon MP, Desmartis M, Lepage-Savary D, Gagnon J, St-Pierre M, Rhainds M, Lemieux R, Gauvin FP, Pollender H, Légaré F. Introducing patients' and the

public's perspectives to health technology assessment: A systematic review of international experiences. Int J Technol Assess Health Care. 2011 Jan;27(1):31-42.

Table 4-1. Benefit-harm categories

	No harms *Patients had no adverse events of interest*	**Minor harms** *Patients only had the adverse events that are less worrying including incident cataracts, prescription-requiring hypertension, prescription-requiring hyperlipidemia, or prescription-requiring infections*	**Moderate harms** *Patients had any of the adverse events that are more worrying including incident glaucoma, requiring cataract surgery and requiring glaucoma surgery*
With benefits *Patients' visual acuity of the better-seeing eye stayed at or improved to 20/40 or better*	**(1) With benefits/No harms**	**(2) With benefits/Minor harms**	**(3) With benefits/Moderate harms**
No benefits *Patients' visual acuity of the better-seeing eye stayed at or decreased to worse than 20/40*	**(4) No benefits/No harms**	**(5) No benefits/Minor harms**	**(6) No benefits/Moderate harms**
		(7) Missing data	

Table 4-2. Data input for benefit-harm assessment

Patient experiences with treatment outcomes over time for the implant and systemic therapy group		
	Implant therapy group	**Systemic therapy group**
	Number of patients, n/N (%)	**Number of patients, n/N (%)**
Visual acuity of the better-seeing eye stayed at or decreased to worse than 20/40 at each time point		
6 months	33/121 (27%)	28/115 (24%)
12 months	37/119 (31%)	24/115 (21%)
18 months	29/114 (25%)	21/117 (18%)
24 months	33/118 (28%)	24/114 (21%)
Number of patients who ever had the event(s) at each time point		
Incident glaucoma		
6 months	0/116 (0%)	0/115 (0%)
12 months	1/116 (1%)	0/115 (0%)
18 months	1/116 (1%)	0/114 (0%)
24 months	27/116 (23%)	7/114 (6%)
Requiring cataract surgery		
6 months	23/124 (19%)	6/121 (5%)
12 months	43/124 (35%)	16/120 (13%)
18 months	65/123 (53%)	25/120 (21%)
24 months	74/121 (61%)	30/119 (25%)
Requiring glaucoma surgery		
6 months	8/124 (6%)	3/121 (2%)
12 months	21/124 (17%)	3/120 (3%)
18 months	33/123 (27%)	4/120 (3%)
24 months	40/121 (33%)	8/119 (7%)
Incident cataracts		
6 months	11/24 (46%)	5/21 (24%)
12 months	22/24 (92%)	8/21 (38%)
18 months	22/24 (92%)	8/21 (38%)
24 months	23/24 (96%)	10/21 (48%)
Prescription-requiring hypertension		
6 months	2/124 (2%)	3/121 (2%)
12 months	4/124 (3%)	4/120 (3%)
18 months	4/123 (3%)	5/120 (4%)
24 months	5/121 (4%)	9/119 (8%)
Prescription-requiring hyperlipidemia		
6 months	1/124 (1%)	3/121 (2%)
12 months	2/124 (2%)	3/120 (3%)
18 months	3/123 (2%)	7/120 (6%)
24 months	3/121 (2%)	8/119 (7%)
Prescription-requiring infections		

6 months	25/124 (20%)	27/121 (22%)
12 months	32/124 (26%)	38/120 (32%)
18 months	38/123 (31%)	52/120 (43%)
24 months	45/122 (37%)	57/119 (48%)

Weights assigned in the main and sensitivity analyses

Outcome	Main analysis	Sensitivity analysis one	Sensitivity analysis two
Visual acuity of the better-seeing eyes stayed at or decreased to worse than 20/40	0.8	1.0	1.0
Incident glaucoma	0.7	1.0	0.5
Requiring cataract surgery	0.6	1.0	0.5
Requiring glaucoma surgery	0.6	1.0	0.5
Incident cataracts	0.3	0.5	0.25
Prescription-requiring hypertension	0.3	0.5	0.25
Prescription-requiring hyperlipidemia	0.3	0.5	0.25
Prescription-requiring infections	0.2	0.5	0.25

Table 4-3. Calculation of the benefit-harm metric at 24 months follow-up

Outcome	Number of cases with the outcome if 1000 patients were treated		Number of cases caused or prevented by implant therapy ($N_X = N_{X,SYS} - N_{X,IMP}$)[1]	Weight assigned to the outcome (W_X)	$N_X * W_X$
	Treated with implant therapy ($N_{X,IMP}$)	Treated with systemic therapy ($N_{X,SYS}$)			
Visual acuity of the better-seeing eyes stayed at or decreased to worse than 20/40	280	211	-69	0.8	-55
Incident glaucoma[2]	226	60	-166	0.7	-116
Requiring cataract surgery	612	252	-359	0.6	-216
Requiring glaucoma surgery	331	67	-263	0.6	-158
Incident cataracts[3]	173	86	-87	0.3	-26
Prescription-requiring hypertension	41	76	34	0.3	10
Prescription-requiring hyperlipidemia	25	67	42	0.3	13
Prescription-requiring infections	369	479	110	0.2	22
Benefit-harm metric (I_X)					**-526**

[1]Numbers of cases caused are negative and numbers of cases prevented are positive.

[2]Assuming 97% of patients at risk at baseline.

[3]Assuming 18% of patients at risk at baseline.

Table 4-4. Main and sensitivity analyses of benefit-harm metrics at different time points after randomization

	Benefit-harm metric	95% confidence interval*	Probability of the metric being positive*
Main analysis			
6 months	-129	-242 to -14	1%
12 months	-317	-436 to -196	0%
18 months	-390	-514 to -264	0%
24 months	-526	-687 to -368	0%
Sensitivity analysis one			
6 months	-201	-362 to -39	1%
12 months	-482	-665 to -298	0%
18 months	-603	-800 to -412	0%
24 months	-808	-1049 to -570	0%
Sensitivity analysis two			
6 months	-115	-245 to 13	4%
12 months	-292	-421 to -161	0%
18 months	-339	-467 to -209	0%
24 months	-439	-588 to -294	0%

*95% confidence interval and the probability of the metric being positive were calculated based on the 10000 bootstrapping replicates.

Table 4-5. Number of patients assigned to each benefit-harm category[1] at 24 months follow-up

		All participants		Stratified by patient's visual acuity of the better-seeing eye at baseline[2]			
				20/40 or better		Worse than 20/40	
		Implant therapy	Systemic therapy	Implant therapy	Systemic therapy	Implant therapy	Systemic therapy
	Sample size	129	126	84	91	44	35
With benefits	No harms	13 (10%)	25 (20%)	11 (13%)	21 (23%)	2 (5%)	4 (11%)
	Minor harms	9 (7%)	36 (29%)	8 (10%)	31 (34%)	1 (2%)	5 (14%)
	Moderate harms	63 (49%)	29 (23%)	50 (60%)	23 (25%)	13 (30%)	6 (17%)
	Total	*85 (66%)*	*90 (71%)*	*69 (82%)*	*75 (82%)*	*16 (36%)*	*15 (43%)*
No benefits	No harms	6 (5%)	12 (10%)	2 (2%)	2 (2%)	3 (7%)	10 (29%)
	Minor harms	6 (5%)	5 (4%)	0 (0%)	2 (2%)	6 (14%)	3 (9%)
	Moderate harms	20 (16%)	7 (6%)	6 (7%)	2 (2%)	14 (32%)	5 (14%)
	Total	*32 (25%)*	*24 (19%)*	*8 (10%)*	*6 (7%)*	*23 (52%)*	*18 (51%)*
Missing data		12 (9%)	12 (10%)	7 (8%)	10 (11%)	5 (11%)	2 (6%)

[1]The definition of benefit and harm categories can be found in the "Methods" section.

[2]One patient in the implant group had missing data on visual acuity at baseline.

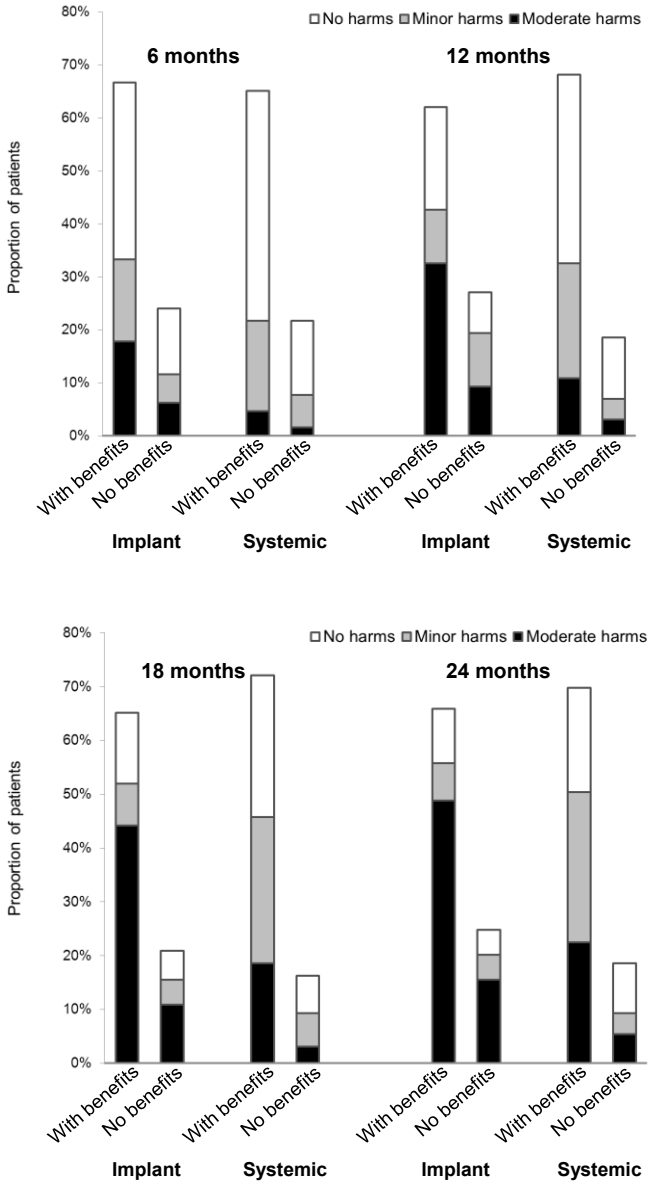


Figure 4-1. Distribution of patients to assigned benefit-harm categories over time by treatment group.

For implant group (n=129) and systemic group (n=126), we categorized the patients first into "With benefits" or "No benefits" and further sub-categorized them into "No harms", "Minor harms" or "Moderate harms". The definition of benefit and harm categories can be found in the "Methods" section. We plotted the distributions over time.

CHAPTER FIVE

CONCLUSIONS

Summary of findings

The goal of treatments is to provide benefits to patients, but they may come with harms. If multiple benefit and harm outcomes are associated with the treatments, it becomes more challenging to make treatment decisions.[1] Doing a comprehensive benefit-harm assessment is thus helpful, even necessary, for different stakeholders, including patients, clinicians, or policy-makers to select the most appropriate treatments. In this dissertation, we conducted a quantitative and patient-centered benefit-harm assessment of corticosteroid implant versus systemic corticosteroids in intermediate, posterior, and panuveitis. Several methodological issues of benefit-harm assessments have also been examined.

Selection of outcomes

Defining the decision-making context is the first and important step of doing a benefit-harm assessment, which involves specifying the population(s), intervention(s), comparison(s), outcome(s), and time frame(s). We emphasize that we should select patient-centered outcomes because this will help generate results that are relevant to patients. Nonetheless, investigators are often challenged by the fact that surrogate outcomes are commonly used in clinical trials to demonstrate treatment effects, instead of patient-centered outcomes.[2] In **chapter 2**, we did a survey of drugs approved by the United States Food and Drug Administration (FDA) to learn how the drug reviewers

dealt with surrogate outcomes when they were assessing the benefits and harms of a drug. We went through the medical review documents of included drug approvals to examine how they considered the evidence of surrogacy. Specifically, we examined if they considered how treatment effects on surrogate outcomes predict treatment effects on patient-centered outcomes.

We were surprised of our findings that the way the FDA dealt with surrogate outcomes was not consistent across disease areas and, that in some diseases, evidence of surrogacy was seldom discussed in the medical reviews. For example in glaucoma, most drugs were only approved based on surrogate outcomes (intraocular pressure), and the rationale for using surrogates was not clearly stated.[3] On the contrary, in diseases such as osteoporosis, FDA required evidence on patient-centered outcomes (fractures) for drug approval.[4] We accordingly proposed a framework for selecting outcomes in a benefit-harm assessment. We emphasize the importance of using patient-centered outcomes when doing a benefit-harm assessment if these outcomes are available. If data on such outcomes are not available to us, we should critically examine the evidence of surrogacy. For example in diabetes, a systematic review of the evidence on how treatment effects on lowering blood sugar levels can predict treatment effects on micro- or macro-vascular events would be extremely helpful.[5] Moreover, we should carefully consider if the evidence of surrogacy can be applied to the specific drugs and populations under review.

Role of patient preferences

Another key element in a benefit-harm assessment is the relative importance of outcomes. To conduct a quantitative benefit-harm assessment in which multiple outcomes are compared against each other, one must properly assign weights to these outcomes in order to reflect their relative importance when estimating the benefit-harm balance.[6] In **chapter 3**, we conducted a survey of patients with uveitis to elicit the relative importance (preferences) of treatment outcomes. We used best-worst scaling (BWS) approach, which is getting more and more popular in health care research,[7] to elicit participants' preferences for the six outcomes that are meaningful to decision-making. In the questionnaire, participants were first asked to read the description of each outcome and complete visual analog scale (VAS) tasks to indicate how serious they perceive the outcomes to be (as a warm-up exercise). Then, they were asked to do the BWS tasks. In each BWS task they made trade-offs between outcomes by indicating the most and least worrying out of three outcomes. Our main findings showed that patients with uveitis considered vision impairment, glaucoma, and need for eye surgery more worrying as compared against cataracts, needing medicine for high blood pressure/cholesterol, or infections.

In our survey, we included two populations of patients with uveitis. One population consisted of patients who were more severe and who had been diagnosed for a longer time period and the other consisted of patients who were less severe and who had

less experiences with the disease. Interestingly, preference results of BWS were not significantly different between these two populations. Furthermore, in regression analyses, we only identified demographic characteristics such as race and educational level to be the factors that may influence their preferences, while other disease characteristics or their prior experiences with outcomes were not significantly associated with the preferences. Another interesting methodological finding of our survey is from comparing the results of BWS versus VAS. We found using BWS approach to elicit patient preferences, a more differentiated distribution of relative importance of outcomes was obtained than using VAS approach. This suggests that using BWS approach seems to be easier than using VAS approach to differentiate the relative importance of outcomes. The preference data we collected provided a basis for assigning weights in our later benefit-harm assessment.

A quantitative benefit-harm assessment based on a comparative effectiveness trial

Results of the Multicenter Uveitis Steroid Treatment (MUST) Trial[8] showed that both corticosteroid implant and systemic corticosteroids work to preserve patient's vision, but it also indicated that these therapies are associated with distinct systemic and ocular adverse effects. We conducted a benefit-harm assessment of these two treatment strategies in patients with intermediate, posterior, and panuveitis in **chapter 4**. We selected the patient-centered outcomes for benefit-harm assessment according to our framework mentioned above. To compare multiple outcomes against each other, we

generated a common comparison metric. We used absolute risk data of the MUST Trial to compute the number of cases prevented or caused by implant therapy as compared to systemic therapy if 1000 patients were treated, and assigned weights to these outcomes (based on our findings of the patient preference survey) to compute the benefit-harm metrics. If the benefit-harm metric was positive, it indicated that implant therapy had a better benefit-harm balance than systemic therapy.

We also examined the joint occurrence of benefit and harm outcomes in the trial participants. Oftentimes, results of each outcome in a clinical trial are reported independently in the literature.[9] But readers have no clues to if the benefit and harm outcomes co-occur in the same group of patients or not. By accounting for the joint occurrence of benefits and harms, we were able to assess their joint impacts on the same patient. We also calculated the benefit-harm metrics at different time points after randomization (up to two years) to learn how the benefit-harm balance would change across time. In addition, we did sensitivity analyses where we assigned different weights to outcomes to evaluate if and how our conclusions may vary. We found the results were consistent across different scenarios and all analyses suggested that systemic therapy may be superior to implant therapy.

Limitations and implications for future research

Here we discuss the limitations of the dissertation and the implications for future research.

In this dissertation, we developed a framework for selecting outcomes in benefit-harm assessments. We emphasize the importance of using patient-centered outcomes whenever possible, but if using surrogate outcomes is inevitable, we should carefully examine the evidence for surrogacy. However, our proposed framework may be limited by that there are insufficient prior studies that have evaluated the validity of surrogate outcomes. The highest level of evidence for surrogacy requires to have randomized clinical trials showing treatment effects on surrogate outcomes predict treatment effects on patient-centered outcomes.[10] Doing such randomized clinical trials with patient-centered outcomes as the primary outcome often needs large sample size and long follow-up time. Limited funding and resources to support a large and long-term study, and lack of interest from industry where most of these drug trials are done would pose real challenges here.[11]

Another issue is that even with good evidence for surrogacy, to make judgments on the generalizability of the evidence can be difficult.[12] For example, there were some large and long-term trials conducted to examine if the treatment effects on lowering blood sugar levels can predict treatment effects on some micro-vascular events.[13, 14]

With this evidence, can we apply it to different drugs in the same class? Probably yes. Can we apply this evidence to the drugs in different classes? Probably not sure since drugs in another class can have an effect on patient-centered outcomes through other causal pathways, and this effect may not be captured by the surrogate outcomes.[15] In the future, we can initiated more large-scale systematic reviews and/or meta-analysis to examine the evidence for surrogacy by different drugs/drug classes. This will allow us to study for what drugs we are more confident in using surrogate outcomes for benefit-harm assessment, and for what drugs there may be knowledge gaps in the evidence for surrogacy. Thus we can focus on these evidence gaps and design studies that address the issue.

Different stakeholders may have different views of the relative importance of outcomes.[16] In our benefit-harm assessment, we decided to adopt a patient-centered approach and incorporated patient preferences in our analysis. At first, we were concerned that patients who had been diagnosed with uveitis for a longer time period may have different preferences as compared to patients who are less experienced with the disease. Some people thought that we should focus on the preferences from patients who are recently diagnosed because their preferences reflect those patients who are in the position to make a treatment decision. However, it was challenging to recruit such patients since it required a large number of clinics to contribute to our study. On the contrary, some people argued that it is probably also necessary to include patients who have more experiences with the disease and treatments because they may be more

familiar with treatment outcomes. Our findings showed that the influences of disease characteristics or their experience with outcomes on preferences, in fact, were limited. Instead, demographic characteristics such as race and education could influence their preferences. The associations we observed were based on our study sample and may not be generalizable to other populations or diseases. Survey of a larger population of patients can be conducted in the future to further examine this issue.

One drawback of this dissertation is that we did not do our benefit-harm assessment prospectively. We, the same as most investigators who plan to do benefit-harm assessments, mainly rely on existing data and our studies are limited by what outcome data have been collected in clinical trials. Most outcome data collected in trials are of interest to clinicians but may not be so to patients. A more ideal way can be, before the trials are initiated, to conduct qualitative studies such as focus groups[17] with relevant patient populations to learn what outcomes are deemed important by patients, and then to elicit the preferences for such outcomes. Not only will this information be useful to those conducting benefit-harm assessments, but also it will be informative to clinical trialists for planning the trial and deciding on what outcome data they need to collect. Before the trial starts, investigators who are doing benefit-harm assessment can pre-specify the outcomes that will be included and pre-specify the weights that are assigned to each outcome in the analysis. This prospective approach can ensure that we don’t miss out the outcomes that are important to patients’ decision-making and also

minimize the bias that may be introduced should the benefit-harm assessment be done *post hoc*.

The issue of defining outcomes for benefit-harm assessment becomes even more important given the preference-elicitation approach we chose for this dissertation. We adopted the BWS approach to assess patient preferences, where we asked participants to make trade-offs among the six outcomes that are meaningful to decision-making. If we included more (or less) outcomes, the preferences results would probably be different, because participants would instead make trade-offs among different sets of outcomes. Also, the benefit and harm outcomes included in our assessment were mainly clinical outcomes. We did not include other important elements such as the burden of the treatments and the cost to patients as we wanted to focus on clinical outcomes first. For patients with uveitis to choose between corticosteroid implant and systemic corticosteroids, these two elements may also be essential to their decision-making because one treatment requires additional surgery of the implant and the other treatment requires long-term use of medications, which incurred distinct burden and cost to patients.[18] Our study only considered clinical outcomes but in the future, we can try adding the other two dimensions into the benefit-harm assessment. It is challenging to do but for sure deserves further study.

Assigning relative importance (weights) to different outcomes is one of the controversial parts of a benefit-harm assessment, because this exercise is by nature

subjective.[19] We did a preference survey of patients with the condition to figure out what the relative importance is for each outcome, but we still find it challenging to transfer these numbers into weights when doing the analysis and to properly communicate with stakeholders about this exercise. Nonetheless, we believe that to do a benefit-harm assessment that is transparent and replicable, documenting the weights used is of great importance and necessary.[20] To overcome this issue, we emphasize the importance of doing multiple sensitivity analyses with different weights assigned to learn if and how the study results would vary.

Our benefit-harm assessment was done from a population perspective. We used preference data from a group of patients with the condition and clinical outcome data of a randomized controlled trial. Thus, we should acknowledge that we are assessing the overall benefit and harms for a population when interpreting our findings. If one wants to apply this evidence to treat an individual, individual patient preferences and their characteristics should thus be considered. Also, the results would be a probabilistic estimate of the treatment being more beneficial than harmful given certain patient characteristics and preferences and would seldom be a yes/no answer. To conduct benefit-harm assessment that is more personalized, we may need to design more large-scale clinical trials (with a more heterogeneous patient group) or conduct modeling studies for benefit-harm assessments under different scenarios.[21] This will help us estimate the benefit-harm balance for patient subgroups with varying preferences and characteristics.

In this dissertation, we examined the health outcomes, factored in the patient preference estimates, and generated a "benefit-harm metric" as a way to summarize the impacts of treatments on patient's health and used it for treatment comparison. An interesting topic for further research may be to see how our approach to benefit-harm assessment can contribute to cost-effectiveness (or cost-utility) analysis. In cost-utility analysis, quality-adjusted life year (QALY) is often the metric chosen to measure the treatment impacts on patient's health, which combines both measures of the length of life and the quality of life.[22] The way to calculate QALY is to multiply the utility value of a health state (e.g., health utility measured by EQ-5D) by the years of life lived in that heath state.[22] However, this approach may be too generic sometimes to capture the impact of treatments on patient's health in many diseases, and it relies on several strong assumptions.[22] Thus, use of QALY for making medical decisions has always been controversial and a recent report from European Consortium in Healthcare Outcomes and Cost-Benefit Research project even recommended against using QALY for healthcare decision-making, since they concluded that QALY is not a scientifically valid estimate.[23] It is yet unclear if our approach, combining trial data on specific clinical outcomes with patient preferences (trade-offs) for these outcomes, would be better or worse as compared to using QALY for making treatment decisions. In the future, studying the advantages and disadvantages of our approach versus using QALY in medical decision-making can be a promising direction for our research.

Mover forward, we believe the patient-centered approach to benefit-harm assessment developed in this dissertation can be further applied to many diseases and settings. A benefit-harm assessment can inform the design, conduct and interpretation of a randomized clinical trial. For instance, when clinical trials are designed to compare treatments with high toxicity in some disease areas such as cancers, investigators may struggle with defining the inclusion/exclusion criteria for trial patients. Concerns that the harms may outweigh the benefits in patients with less severe diseases often make it difficult to justify the inclusion for such patients. Therefore, using benefit-harm assessments to model the benefit-harm balance in treating these patients would help investigators deal with the issue.[24] During the conduct of a clinical trial, one major task for investigators as well as for Data and Safety Monitoring Committee (DSMC) is to monitor treatment efficacy and its safety. In instances where treatments may cause various outcomes that occur with different frequency and have different level of severity, a quantitative benefit-harm assessment will be helpful to synthesize data and provides a guidance for investigators and the DSMC as to whether the trial should be continued or should be terminated early for benefits (or for harms).[25] Finally, to interpret the finding from a clinical trial, benefit-harm assessment would be valuable as shown by the studies conducted in this dissertation. Benefit-harm assessment also plays a role in translating evidence generated by clinical trials into clinical practice. It is closely related to the development of risk-stratified treatment recommendations by guideline developers since their objective is to identify those patients who will have most favorable treatment benefit-harm balance.[26]

To conclude, in this dissertation we have demonstrated a framework for doing patient-centered benefit-harm assessment, which involves selecting patient-centered outcomes, eliciting patient preferences, and integrating patients' perspective into clinical trial data on benefits and harms. The applications of our approach will help patients make evidence- and preference-based treatment decisions. Our goal is to make medicine more safe and effective and to improve public health.

REFERENCES

1. Guo JJ, Pandey S, Doyle J, Bian B, Lis Y, Raisch DW. A review of quantitative risk-benefit methodologies for assessing drug safety and efficacy-report of the ISPOR risk-benefit management working group. Value Health. 2010;13(5):657-66.

2. Gandhi GY, Murad MH, Fujiyoshi A, Mullan RJ, Flynn DN, Elamin MB, et al. Patient-important outcomes in registered diabetes trials. JAMA. 2008 Jun 4;299(21):2543-9.

3. Weinreb RN, Kaufman PL. The glaucoma research community and FDA look to the future: A report from the NEI/FDA CDER glaucoma clinical trial design and endpoints symposium. Invest Ophthalmol Vis Sci. 2009 Apr;50(4):1497-505.

4. The Food and Drug Administration. Guidelines for preclinical and clinical evaluation of agents used in the prevention or treatment of postmenopausal osteoporosis. http://www.fda.gov/downloads/ScienceResearch/SpecialTopics/WomensHealthResearch/UCM131206.pdf. 1994 (last accessed on June 10, 2014).

5. Fleming TR, Powers JH. Biomarkers and surrogate endpoints in clinical trials. Stat Med. 2012 Nov 10;31(25):2973-84.

6. Puhan MA, Singh S, Weiss CO, Varadhan R, Boyd CM. A framework for organizing and selecting quantitative approaches for benefit-harm assessment. BMC Med Res Methodol. 2012 Nov 19;12:173,2288-12-173.

7. Louviere JJ, Lings I, Islam T, Gudergan S, Flynn T. An introduction to the application of (case 1) best–worst scaling in marketing research. Int J Res Mark. 2013;30(3):292–303.

8. Multicenter Uveitis Steroid Treatment (MUST) Trial Research Group, Kempen JH, Altaweel MM, Holbrook JT, Jabs DA, Louis TA, Sugar EA, Thorne JE. Randomized comparison of systemic anti-inflammatory therapy versus fluocinolone acetonide implant for intermediate, posterior, and panuveitis: the multicenter uveitis steroid treatment trial. Ophthalmology. 2011;118(10):1916-26.

9. Kraemer HC, Frank E. Evaluation of comparative treatment trials: assessing clinical benefits and risks for patients, rather than statistical effects on measures. JAMA. 2010;304(6):683-4.

10. International Conference on Harmonization. ICH harmonized tripartite guideline. Statistical principles for clinical trials - E9. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf. 1998 (last accessed on February 1, 2014).

11. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. Clin Pharmacol Ther. 2001 Mar;69(3):89-95.

12. Baker SG, Kramer BS. Surrogate endpoint analysis: an exercise in extrapolation. J Natl Cancer Inst. 2013 Mar 6;105(5):316-20. doi: 10.1093/jnci/djs527. Epub 2012 Dec 21.

13. American Diabetes Association. Standards of medical care in diabetes--2013. Diabetes Care. 2013 Jan;36 Suppl 1:S11-66.
14. Duckworth W, Abraira C, Moritz T, Reda D, Emanuele N, Reaven PD, et al. Glucose control and vascular complications in veterans with type 2 diabetes. N Engl J Med. 2009 Jan 8;360(2):129-39.
15. Svensson S, Menkes DB, Lexchin J. Surrogate outcomes in clinical trials: a cautionary tale. JAMA Intern Med. 2013 Apr 22;173(8):611-2.
16. van Til JA, Ijzerman MJ. Why should regulators consider using patient preferences in benefit-risk assessment? Pharmacoeconomics. 2014 Jan;32(1):1-4.
17. von Wagner C1, Halligan S, Atkin WS, Lilford RJ, Morton D, Wardle J. Choosing between CT colonography and colonoscopy in the diagnostic context: a qualitative study of influences on patient preferences. Health Expect. 2009 Mar;12(1):18-26.
18. Multicenter Uveitis Steroid Treatment Trial Research Group, Kempen JH, Altaweel MM, Holbrook JT, Jabs DA, Sugar EA. The multicenter uveitis steroid treatment trial: rationale, design, and baseline characteristics. Am J Ophthalmol. 2010;149(4):550-561.e10.
19. The Food and Drug Administration. Structured approach to benefit-risk assessment in drug regulatory decision-making. Draft PDUFA V implementation plan - February 2013 fiscal years 2013-2017. http://www.fda.gov/downloads/ForIndustry/UserFees/PrescriptionDrugUserFee/UCM329758.pdf. 2013 (last accessed on June 10, 2014).

20. Chuang-Stein C. Quantitative risk/benefit assessment: where are we? Proceedings of the Fourth Seattle Symposium in Biostatistics: Clinical Trials. Springer New York. 2013;119-135.

21. Gail MH. Using absolute risks to assess the risks and benefits of treatment. Thorax. 2014. [Epub ahead of print].

22. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the economic evaluation of health care programmes. Oxford university press, 2005.

23. European Consortium in Healthcare Outcomes and Cost-Benefit Research. European guidelines for cost-effectiveness assessments of health technologies. http://www.echoutcome.eu/images/Echoutcome__Leaflet_Guidelines___final.pdf. 2013 (last accessed on August 9, 2014).

24. Slankamenac K. Complications in abdominal surgery: Assessment, prediction and prevention (PhD thesis). University of Amsterdam, the Netherlands. http://dare.uva.nl/record/465611. 2014 (last accessed on August 12, 2014).

25. Freedman L, Anderson G, Kipnis V, Prentice R, Wang CY, Rossouw J, Wittes J, DeMets D. Approaches to monitoring the results of long-term disease prevention trials: examples from the Women's Health Initiative. Control Clin Trials. 1996 Dec; 17(6): 509-525.

26. Yu T, Vollenweider D, Varadhan R, Li T, Boyd C, Puhan MA. Support of personalized medicine through risk-stratified treatment recommendations - an environmental scan of clinical practice guidelines. BMC Med. 2013 Jan 9;11:7.

CURRICULUM VITAE

TSUNG YU

PERSONAL DATA

PhD candidate

Department of Epidemiology

Johns Hopkins Bloomberg School of Public Health

615 North Wolfe Street, Baltimore, MD 21205, USA

Email: tsyu@jhsph.edu

Date of Birth: January 20, 1986

Place of Birth: Tainan, Taiwan

RESEARCH INTERESTS

- Clinical Trials Methodology
- Systematic Reviews and Meta-Analysis
- Benefit-Harm Assessment
- Patient Preferences
- Evidence-Based Health Care
- Clinical Epidemiology

EDUCATION

2014 (expected)	**Johns Hopkins Bloomberg School of Public Health**, Baltimore, MD, USA
	PhD, Epidemiology
	Concentration: Clinical Trials and Evidence Synthesis
	Advisors: Professors Milo A. Puhan and Janet T. Holbrook
2011	**Johns Hopkins Bloomberg School of Public Health**, Baltimore, MD, USA

ScM, Epidemiology

Advisor: Professor Milo A. Puhan

2008 **National Taiwan University**, Taipei, Taiwan

BS, Life Science

AWARDS

2013 Doctoral Thesis Research Fund, Johns Hopkins Bloomberg School of Public Health

2012-2013 Scholarship of Government Sponsorship for Overseas Study, Taiwan

2011-2012 Cochrane Eyes and Vision Group Fellowship, Johns Hopkins Bloomberg School of Public Health

2011-2012 Student Research and Travel Award, Center for Clinical Trials, Johns Hopkins University

2010 Masters Tuition Scholarship, Johns Hopkins Bloomberg School of Public Health

2008 Dean Award, College of Life Science, National Taiwan University

2007 National Science Council Research Scholarship, Taiwan

RESEARCH EXPERIENCE

2013-Present **Graduate Research Assistant**

MUST and ALA-ACRC Data Coordinating Center, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

MUST (Multicenter Uveitis Steroid Treatment Trial); ALA-ACRC (American Lung Association Asthma Clinical Research Center)

2010-Present **Graduate Research Assistant**

Cochrane Eyes and Vision Group, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

2006-2008 **Research Assistant**

Cell Biology and Morphology Lab, Institute of Molecular and Cellular Biology, National Taiwan University, Taipei, Taiwan

EDITORIAL ACTIVITIES

Peer Review Activities

- *Ophthalmology*
- *British Journal of Ophthalmology*

PUBLICATIONS

Peer Reviewed Journal Articles

Published

Juo YY, Johnston F, Zhang D, Juo HH, Wang H, Pappou EP, **Yu T**, Easwaran H, Baylin S, Engeland M, Ahuja N. Prognostic value of CpG island methylator phenotype among colorectal cancer patients: a systematic review and meta-analysis. **Ann Oncol**. 2014 Apr 8. [Epub ahead of print]

Li T, Saldanha JI, Vedula SS, **Yu T**, Rosman L, Twose C, Goodman S, Dickersin K. Learning by doing – teaching systematic review methods in 8 weeks. **Research Synthesis Methods**. 2014. [Epub ahead of print]

Yu T, Fain K, Boyd CM, Singh S, Weiss CO, Li T, Varadhan R, Puhan MA. Benefits and harms of roflumilast in moderate to severe COPD. **Thorax**. 2013 Dec 17. [Epub ahead of print]

Ter Riet G, Chesley P, Gross AG, Siebeling L, Muggensturm P, Heller N, Umbehr M, Vollenweider D, **Yu T**, Akl EA, Brewster L, Dekkers OM, Mühlhauser I, Richter B, Singh S, Goodman S, Puhan MA. All that glitters isn't gold: a survey on acknowledgment of limitations in biomedical studies. **PLoS One**. 2013 Nov 20;8(11):e73623.

Yu T, Li T, Lee KJ, Friedman DS, Dickersin K, Puhan MA. Setting priorities for comparative effectiveness research on management of primary angle closure: a survey of Asia-Pacific clinicians. **J Glaucoma**. 2013 Jul 3. [Epub ahead of print]

Yu T, Vollenweider D, Varadhan R, Li T, Boyd C, Puhan MA. Support of personalized medicine through risk-stratified treatment recommendations - an environmental scan of clinical practice guidelines. **BMC Med**. 2013 Jan 9;11:7.

In preparation

Yu T, Hsu YJ, Fain KM, Boyd CM, Holbrook JT, Puhan MA. Use of surrogate outcomes in FDA drug approvals. (dissertation chapter 2)

Yu, et al. Patients with non-infectious uveitis consider impaired vision, glaucoma and need for eye surgery worrying adverse treatment outcomes: results of a best-worst scaling study (dissertation chapter 3).

Yu, et al. Using a patient-centered approach to benefit-harm assessment in treatment decision-making: a case study of corticosteroid therapy in uveitis (dissertation chapter 4).

Stegeman I, Bossuyt PM, Yu T, Boyd C, Puhan MA. Aspirin for primary prevention of cardiovascular disease and cancer: a benefit and harm analysis.

Fain, et al. Evidence selection in a prescription drug's benefit-harm assessment.

Li, et al. Uses and misuses of crossover trials in evaluating medical interventions for glaucoma.

PRESENTATIONS

Oral Presentation

Juo YY, Johnston F, Zhang D, Juo HH, Wang H, Emmanouil P, **Yu T**, Easwaran H, Baylin S, Ahuja N. Prognostic value of CpG island methylator phenotype (CIMP) among colorectal cancer patients: a systematic review and meta-analysis. **The American College of Surgeons 2014 Clinical Congress**, San Francisco, October, 2014

Yu T, Li T, Hawkins B, Dickersin K. Reporting of crossover trials on medical interventions for glaucoma. **The Seventh International Congress on Peer Review and Biomedical Publication**, Chicago, USA , September, 2013

Poster Presentations

Wang X, Hui X, Lindsley K, Li T, **Yu T**, Wormald R, Dickersin K. Priority setting project for open-angle glaucoma and angle-closure glaucoma: where we are now. **The 22nd Cochrane Colloquium**, Hyderabad, India, September, 2014

Yu T, Fain KM, Singh S, Weiss CO, Li T, Varadhan R, Boyd CM, Puhan MA. Comparison of benefits and harms of roflumilast in patients with moderate to severe chronic obstructive pulmonary disease and a history of exacerbations. **Society for Clinical Trials Annual Meeting**, Boston, USA , May, 2013

Saldanha IJ, Vedula SS, **Yu T**, Rosman L, Twose C, Li T, Dickersin K. Learning by doing – teaching systematic review methods in 8 weeks. **The 20th Cochrane Colloquium**, Auckland, New Zealand, October, 2012

Yu T, Li T, Dickersin K. Patient-centered outcome reporting in clinical trials and systematic reviews on medical interventions for primary open-angle glaucoma. **The 20th Cochrane Colloquium**, Auckland, New Zealand, October, 2012

Yu T, Li T, Puhan M, Dickersin K. Setting priorities for comparative effectiveness research on the management of primary angle closure: a survey of Asia-Pacific clinicians. **The 19th Cochrane Colloquium**, Madrid, Spain, October, 2011

TEACHING EXPERIENCE

Teaching Assistant

2014	Comparative Effectiveness Research: Outcome Measurement, Johns Hopkins Bloomberg School of Public Health
2014	Clinical Trials Management, Johns Hopkins Bloomberg School of Public Health
2012-2014	Systematic Review and Meta-analysis, Johns Hopkins Bloomberg School of Public Health
2012	Introduction to Clinical Trials, Johns Hopkins Bloomberg School of Public Health
2011	Epidemiologic Methods II, Johns Hopkins Bloomberg School of Public Health
2011	Diagnosis, Prediction and Screening Methods in Clinical Research, Johns Hopkins Bloomberg School of Public Health
2010-2011	Principles of Epidemiology, Johns Hopkins Bloomberg School of Public Health
2008	Biotechnology Core Techniques, National Taiwan University

Clinical Trials Journal Club Co-coordinator

2013-2014	Center for Clinical Trials, Johns Hopkins University
2011-2012	Center for Clinical Trials, Johns Hopkins University

PROFESSIONAL DEVELOPMENT

Language Skills: English (Fluent), Mandarin (Fluent)

Computer Skills: R, SAS, STATA, C#, Microsoft Office Suite